

字段格式

qq 号, 性别, 出生年份, 上网场景, 婚姻状态, 学历, 专业, 广告创意 id, 广告类目 id, 广告系列 id, 广告主 id, 广告商品类型, 广告商品 id, 图片 url, 落地页 url, 标题及描述, 曝光时间, 广告位 id, 点击数量

```
B = foreach B generate
  A::neg_data::qq_md5,
  qq_demographics::gender,
  qq_demographics::year,
  qq_demographics::surf_scene,
  qq_demographics::marriage_status,
  qq_demographics::education,
  qq_demographics::profession,

  A::neg_data::creative_id,
  A::neg_data::category_id,
  A::neg_data::series_id,
  A::neg_data::advertiser_id,
  A::neg_data::product_type,
  A::neg_data::product_id,
  A::creative_info::img_url,
  A::creative_info::dest_url,
  A::creative_info::literal,

  A::neg_data::imp_time,
  A::neg_data::pos_id,

  A::neg_data::click_num;
```

数据示例

000031aff07c058149a44cfcc713dee8 1 1993 2 1 4 1 20326581 40512 1266319 479355
19 417200582
http://pgdt.gtimg.cn/gdt_crt_65/0/DAAB1B7APoAlwABXBW1QBABrqb7UGz.jpg/0?ck=26e8c070868700e12b05a4d9fe266d94 <https://itunes.apple.com/cn/app/wei-pin-hui-chun-shang-xin/id417200582?mt=8&uo=4> 不会搭配的男生进来看, 不用一百元就能变得帅帅的! 1458176164 288235913187546319 1

00004a4fdde2df45cd85698f72a44643 2 1985 -1 -1 -1 -1 19477387 801 595816 24439 3
1239629206
http://pgdt.gtimg.cn/gdt_crt_65/0/CAAFA93BBBWLfC3Atb4QQb-.jpg/0?ck=8aaebab7bbc5de818f441ff9ce4ab756 <http://user.qqzone.qq.com/1239629206> 早晨, 直接用清水洗脸, 越洗越黑, 点关注, 这样洗, 白过牛奶 1458202337 288235913187546319 0

各个字段解释

qq 号, 性别, 出生年份, 上网场景, 婚姻状态, 学历, 专业, 广告创意 id, 广告类目 id, 广告系列 id, 广告主 id, 广告商品类型, 广告商品 id, 图片 url, 落地页 url, 标题及描述, 曝光时间, 广告位 id, 点击数量

【qq 号】 qq_md5:

经过 md5 加密之后的 qq 号

【性别】 gender

【出生年份】 year

【上网场景】 surf_scene

【婚姻状态】 marriage_status

【学历】 education

【专业】 profession

以上除了“出生年份”外均为枚举值, 含义见附件中的 *user_data.proto* 文件

【广告创意 id】 creative_id

【广告类目 id】 category_id

广告类目和广告内容的语义信息有关

具体值的含义在附件“*广告主行业准入规则.xlsx*”中有具体描述

【广告系列 id】 series_id

【广告主 id】 advertiser_id

广告有树状层级结构, 自顶向下为: 广告主 - 广告系列 - 广告商品 - 广告创意

每个层级的节点都有 id, 最细粒度的是广告创意 id

【广告商品 id】 product_id

【广告商品类型】 product_type

在层级结构上对应“广告商品”, 每个广告商品对应着 0 个或 1 个落地页 (点击后跳转的页面)

如果没有对应的落地页, 则广告商品 id 为空

广告商品类型和广告来源相关, 值代表的具体类型参见 *producttype.png*

【图片 url】

通过此 url 可以下载到该广告图片

【落地页 url】

落地页为点击之后转到的页面

【标题及描述】

一般是中文字符串

【曝光时间】 imp_time:

Unix 时间

【广告位 id】 pos_id:

只有四种广告位 id

【点击数量】 click_num

0 或者 1

0 表示负反馈，表示用户对该广告不感兴趣

1 表示正反馈，表示用户点击了该广告