|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **data变量数据分布表（11064800条）** | | | | **数据范围** | **Label为1时的**  **平均值** | **Label为1时的**  **众数** |
| 1 | aid | 广告 ID |  | 6~2216 | 1044.576105 | 692 |
| 2 | uid | 用户 ID |  | 2~82542900 | 41275210 | 27479691 |
| 3 | label | 是否合适 |  | 0（不合适,数量最多）  1（合适）  -1(未知) |  |  |
| 4 | advertiserId | 广告主ID | 账户结构分为四级：账户、推广计划、广告、素材 | 0~78 | 33.041644 | 26 |
| 5 | campaignId | 推广计划 ID |  | 0~137 | 65.996644 | 109 |
| 6 | creativeId | 素材 ID | 一条广告下可以有  多组素材（所以ID代表素材数量？） | 0~172 | 81.291717 | 37 |
| 7 | creativeSize | 素材大小 |  | 20~109 | 53.228872 | 35 |
| 8 | adCategoryId | 广告类目 |  | 0~39 (10以下居多) | 11.448752 | 6 |
| 9 | productId | 商品 ID |  | 0~32 （0最多） | 6.45308 | 0 |
| 10 | productType | 商品类型 | 如京东 -- 商品、  app-- 下载 | 0~3 | 1.389204 | 0 |
| 11 | age | 年龄 | 每个序号表示一个年龄分段 | 0~5（0相当多） | 2.781115 | 2 |
| 12 | gender | 性别 | 男 / 女 | 0~2（0相当多） | 1.320577 | 1 |
| 13 | marriageStatus | 婚姻状况 | 单身 / 已婚等状态（多个状态可共存） | 0~25 ，多值 | 3.212731 | 1 |
| 14 | education | 学历 | 博士 / 硕士 / 本科 / 高中 / 初中 / 小学 | 0~7（0相当多） | 4.211456 | 7 |
| 15 | consumptionAbility | 消费能力 | 高 / 低 | 0~2（0相当多） | 1.055401 | 1 |
| 16 | LBS | 地理位置 | 每个序号代表一个地理位置 | 0~855（0相当多） | 379.148163 | 387 |
| 17 | interest1 | 兴趣特征组1 | 每个兴趣特征组包  含若干个兴趣 ID |  |  |  |
| 18 | interest2 | 兴趣特征组2 |  |  |  |
| 19 | interest3 | 兴趣特征组3 |  |  |  |
| 20 | interest4 | 兴趣特征组4 |  |  |  |
| 21 | interest5 | 兴趣特征组5 |  |  |  |
| 22 | kw1 | 关键词特征组1 | 较兴趣类目更细粒度地表示用户喜好，每个关键词特征组包含若干用户感兴趣的关键词 |  |  |  |
| 23 | kw2 | 关键词特征组2 |  |  |  |
| 24 | kw3 | 关键词特征组3 |  |  |  |
| 25 | topic1 | 用户喜好主题1 | 使用 LDA 挖掘的用户喜好主题 |  |  |  |
| 26 | topic2 | 用户喜好主题2 |  |  |  |
| 27 | topic3 | 用户喜好主题3 |  |  |  |
| 28 | appIdInstall | APP近期安装行为 |  |  |  |  |
| 29 | appIdAction | APP 活跃 | 用户使用的活跃 APP |  |  |  |
| 30 | ct | 上网连接类型 | WIFI/2G/3G/4G | 0~4 ，多值 | 21.626224 | 34 |
| 31 | os | 操作系统 | Android/IOS | 0~4 ，多值 | 1.280066 | 1 |
| 32 | carrier | 移动运营商 | 移动 / 联通 / 电信 / 其他 | 0~3（0相当多） | 1.056752 | 1 |
| 33 | house | 有房 | 是否有房 | 0,1（0相当多） | 0.165608 | 0 |