目前为了评估广告效果，希望基于广告曝光数据进行分析（广告曝光用户数据集）：

1. 在全量数据及（data\_all.csv）中找到与广告曝光用户相似的用户，并看用户在不同特征上的分布差异？

PSM匹配方法：https://zhuanlan.zhihu.com/p/559411557

1. 广告曝光用户人均消费是否高于未被广告曝光且相似用户？在哪个行为类型（behavior\_type）的专户上更高？

数据集说明

表1：data\_all（全量用户数据，包含广告曝光用户）

| 字段 | 字段说明 | 提取说明 |
| --- | --- | --- |
| user\_id | 用户标识 | 抽样&字段脱敏 |
| item\_id | 商品标识 | 字段脱敏 |
| behavior\_type | 用户对商品的行为类型 | 包括浏览、收藏、加购物车、购买，对应取值分别是1、2、3、4。 |
| item\_category | 商品分类标识 | 字段脱敏 |
| revenue | 收入 | 用户购买金额 |

表2：广告曝光用户（同上）

表3：user\_type

| 字段 | 字段说明 | 提取说明 |
| --- | --- | --- |
| user\_id | 用户标识 | 抽样&字段脱敏 |
| gender | 用户性别 | 0为女性；1为女性 |
| Income | 用户收入水平编码 |  |
| Region | 用户地区编码 |  |
| Age | 用户年龄编码 |  |
| is\_member | 会员 | 0为非会员；1为会员 |