**数据挖掘目标**：根据用户历史的充值、购买、消费记录等信息，预测其在未来一个半月内是否会激活为医美用户（预测环节），以及特征（答辩环节）

**数据文件描述**：所有的X\_开头文件都是可以利用的消费者历史信息，Y\_submit.xlsx是最后提交的预测文件格式，选手需要对每一个客户ID的激活概率进行预测，并按Y\_submit.xlsx的格式（不能打乱用户顺序）提交各用户被激活为医美用户的“**概率**”

**预测评价标准**：提交截止后，裁判会根据选手提交的预测概率Y\_submit.xlsx与真实的激活情况进行对比，以AUC指标为优化目标进行评估，计算方法参见：

<https://baike.baidu.com/item/AUC/19282953?fr=aladdin>

例如，可以使用python包sklearn的metrics.roc\_auc\_score进行计算

**数据发布**：

数据发放时间：2021年7月18日早上9:00 AM

数据发放平台：

1. 比赛官方微信群
2. 百度网盘链接：https://pan.baidu.com/s/1YlmiwpR0Y-eXXlUpUvZH\_A 提取码：e7om

**预测结果提交**：

允许提交次数：3次，超过3次提交的成绩作废

预测AUC得分：取3次提交中最高分

AUC反馈方式：提交截止前每天中午12:00 am更新即时AUC得分和排名

提交渠道为邮件发送submit.xlsx至邮箱：harveydeng.ms@gmail.com

邮件名为“用户转化竞赛-*队长姓名-*提交次数”

例如：用户转化竞赛-张三*-*1

提交截止时间：2021年7月23日 下午5:00 pm

比赛要求：只能由队伍成员进行分析讨论并完成编程，不得请求任何其他非队伍成员参与。我们将在最终答辩中评估各位同学

预测排名分 =

**答辩环节**：2021年7月25日 9:00，复旦大学光华楼西辅楼106

**答辩内容**：阐述小组对问题的分析过程与结果，并重点展示对“特征”的挖掘，回答企业最想知道的内容：什么样的客户（特征）容易被激活为医美用户。

答辩评分：阐述的清晰度，方法的创新度，特征挖掘的价值度等，

**最终评分权重**：预测排名分(10分制) \* 0.5 + 答辩得分（10分制） \* 0.5