场景 A

首先对行为表训练集进行处理，过滤掉缺失值超过 70%的变量，将缺失值单独另设值，将数值型变量的缺失值设为-1，离散型设为‘novalue’，在过滤掉包含缺失值只有两个取值的变量，对 target 进行连接，提取每个变量的 iv 编码值，过滤掉 iv<0.02 的值。

对消费表进行处理，根据 RPM 消费行为模型，衍生新的特征，即消费的频次，频率，最后消费时间距离分析基准时间的天数，提取消费总金额，平均消费金额，以及对几个重要度较高的变量进行衍生，包括每个 id 的 v2.v3 出现频率最高的取值，每个用户 V10 变量的总和和平均值，最后提取每个用户在每个季节的消费额以提取用户消费行为的季节特点。

对于查询表，衍生最后一次消费距分析时间天数的变量，以及查询出现的频次，最后将原始表格的分类变量转化哑变量形式，根据用户 id，求这些哑变量之和，含义为某用户的某个变量，查询出现某个值的次数。有些用户 id 不再查询表中，则全部用 0 值进行插补。

最后将三个表连接起来，使用 Xgboost 模型，Lightgbm 模型，Gradient boost 模型进行加权平均模型融合。

场景 B

根据推断场景 B 和 A 的样本均来自一个总体，所以利用 A 场景的行为表和消费表，衍生特征后，采用和场景 A 同样过程建模后，预测场景 B

最后一次提交的结果:

相比上一次放宽了变量过滤的条件，采用更多的行为表变量，场景 A 的 auc 提升，场景 b 的 auc 下降。