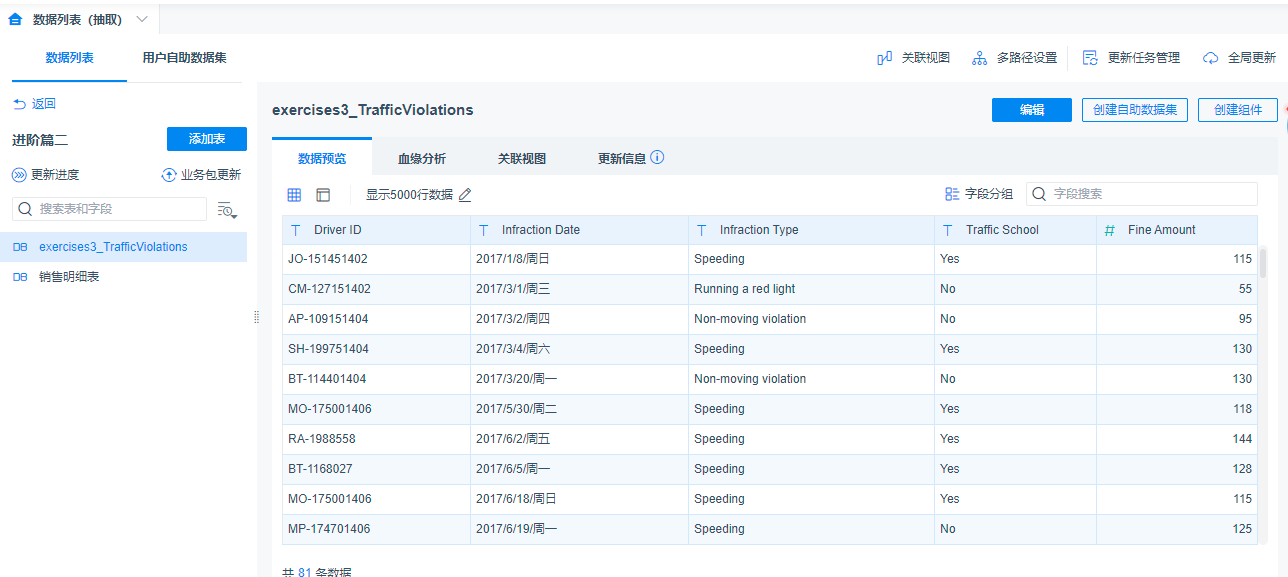
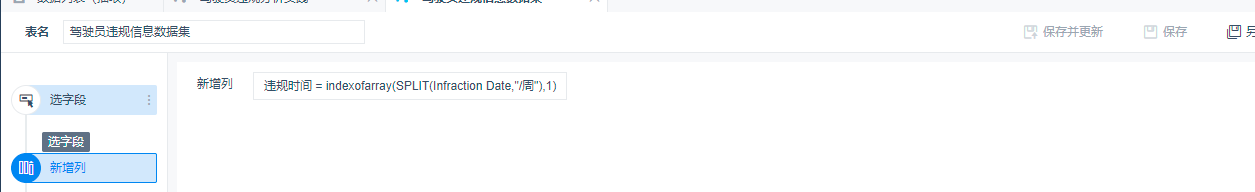
驾驶员最早两次违规时间差分析实践

一、上传数据源

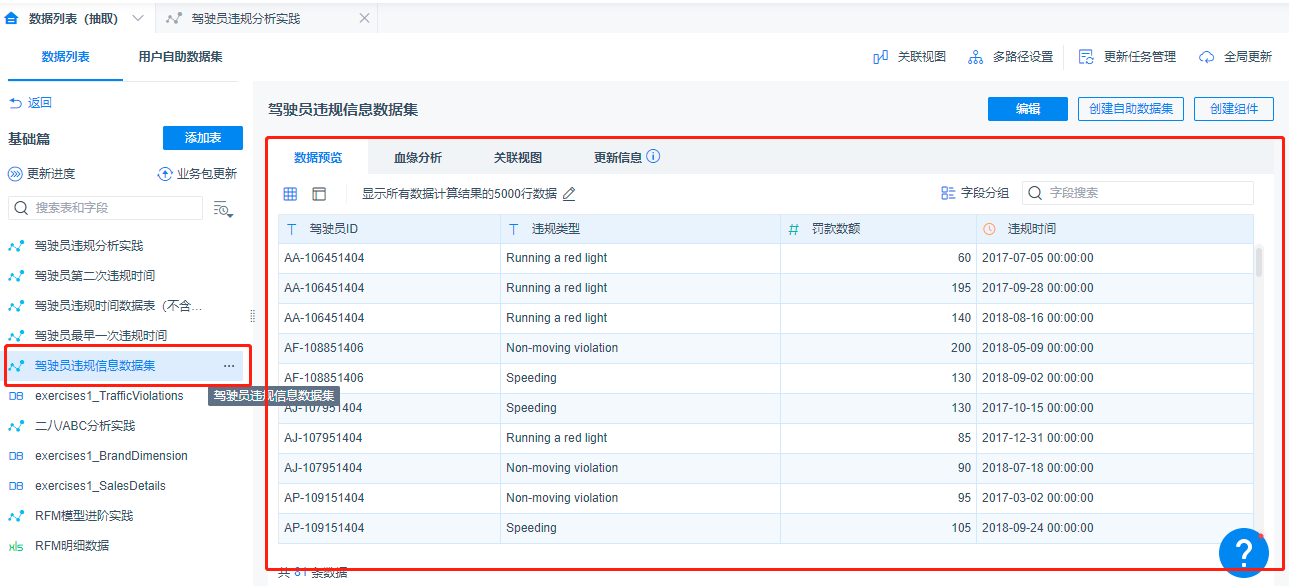


二、创建自助数据集

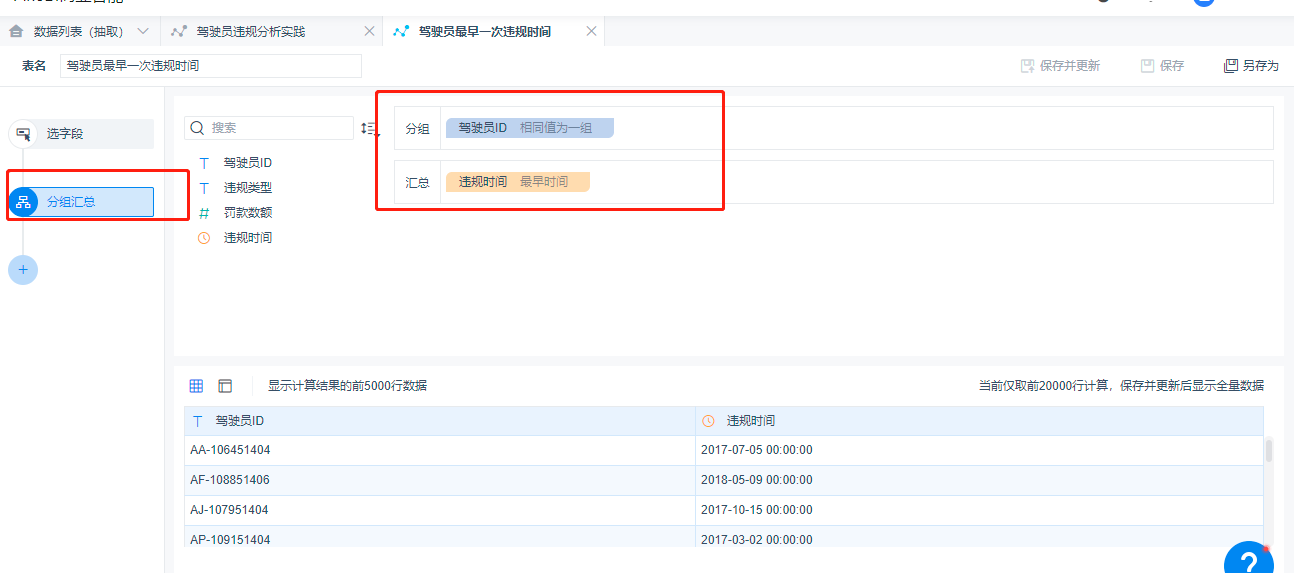
1）驾驶员违规记录表：将数据整理成合理字段样式，选取“驾驶员ID”，“违规类型”，“罚款数额”，“违规时间”四个字段进行分析，将日期字段分割改为时间字段







2）驾驶员最早一次违规时间：

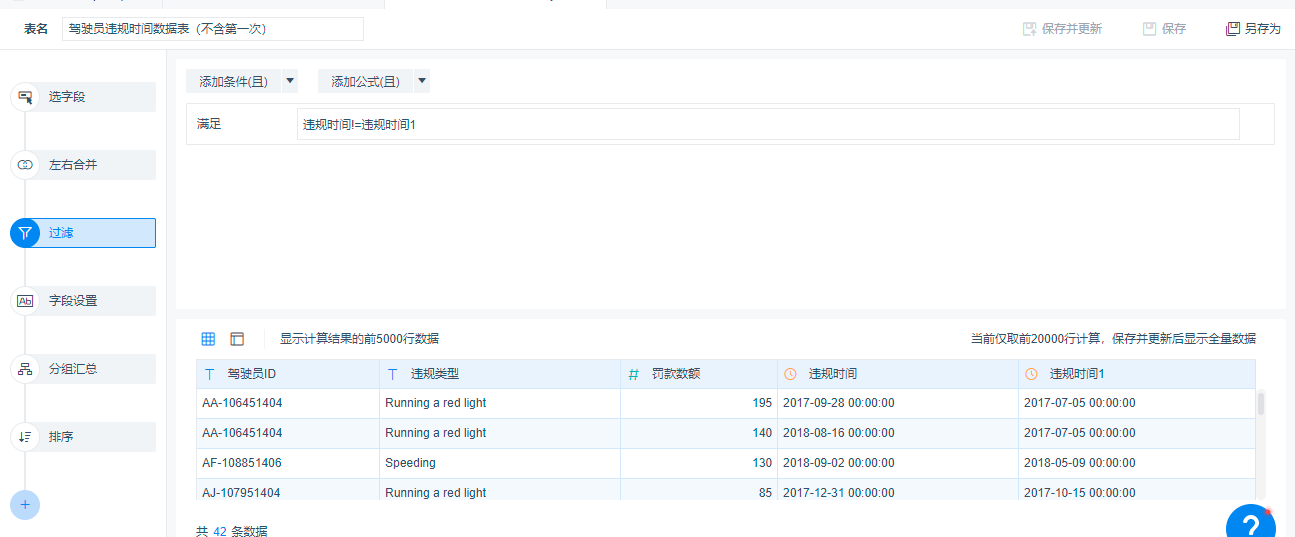


3）驾驶员违规时间数据表（不含第一次）

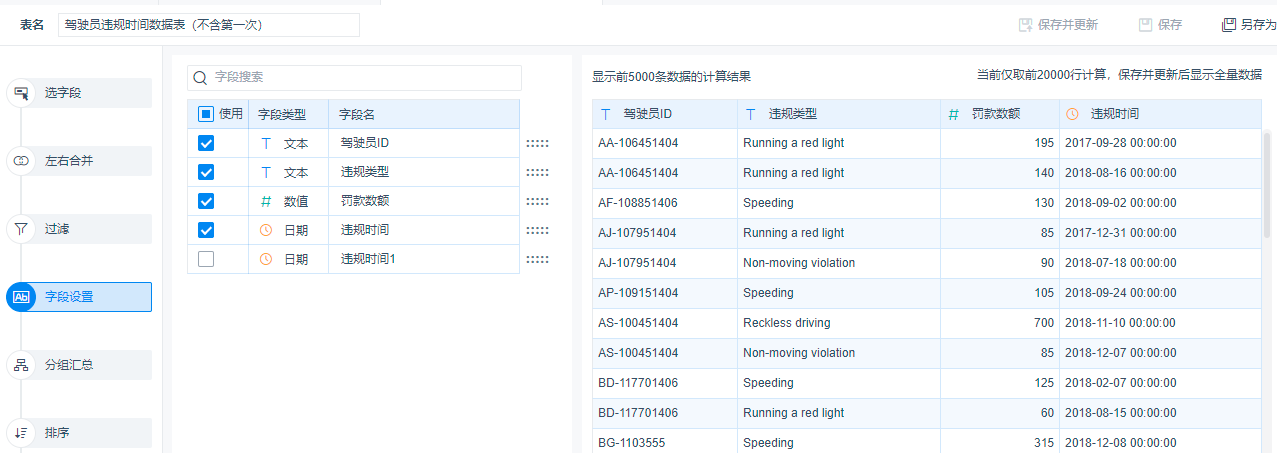
①选取驾驶员违规信息数据表的数据，与“驾驶员最早一次违规时间”进行左合并，根据驾驶员ID匹配第一次违规时间

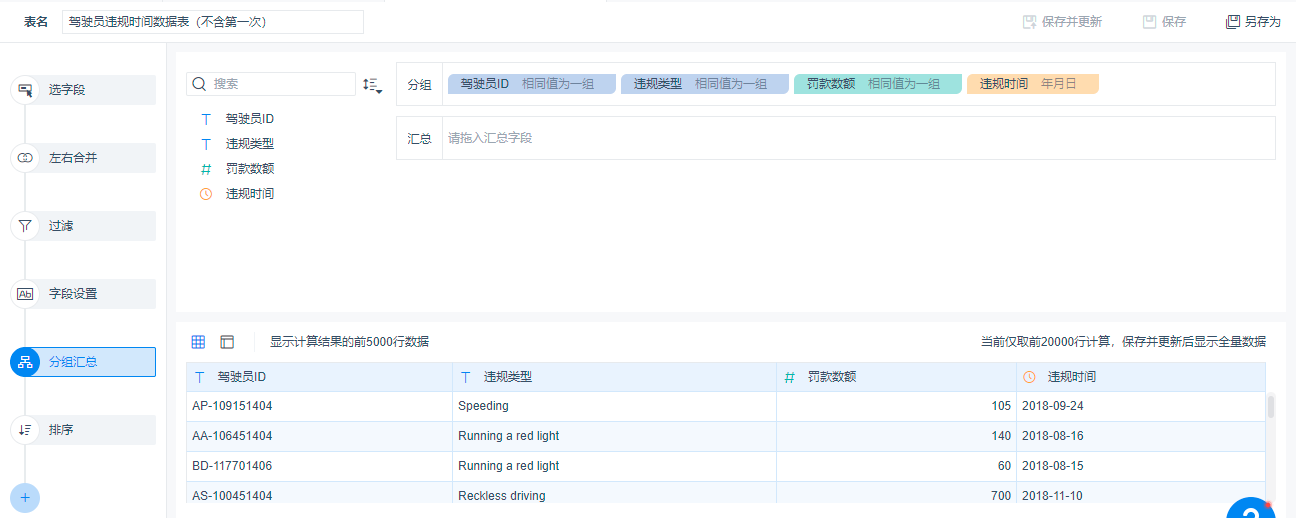


②过滤：将“违规时间”与“违规时间1”相同的过滤掉，即去掉总表中每个驾驶员的第一次违规的数据。

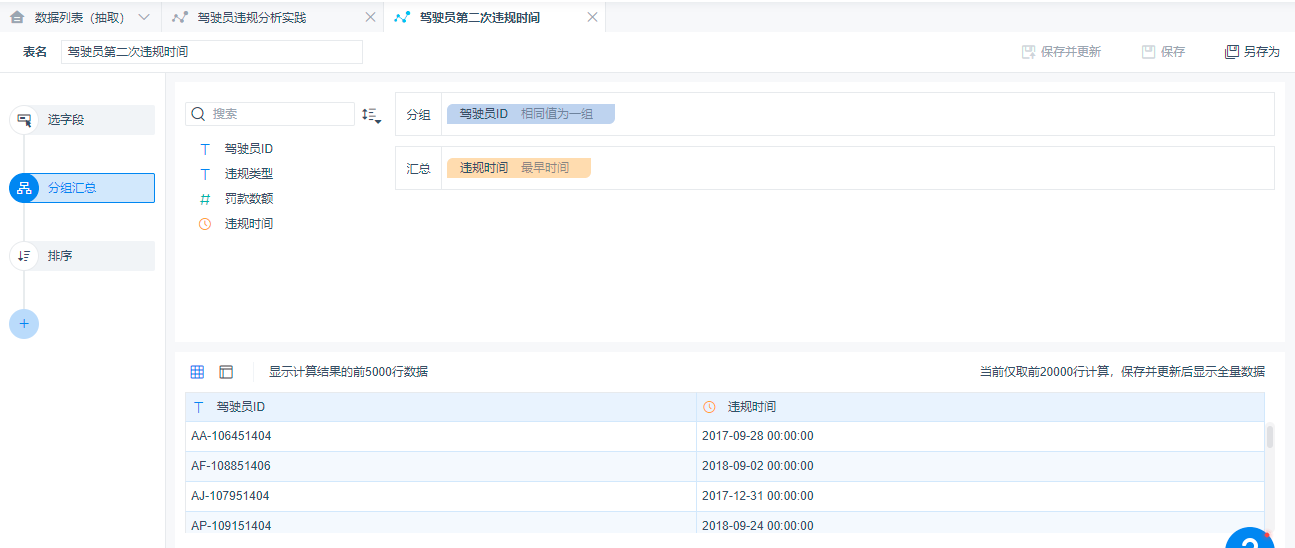


③字段设置；分组汇总，隐藏“违规时间1”，并利用分组汇总去掉重复数据





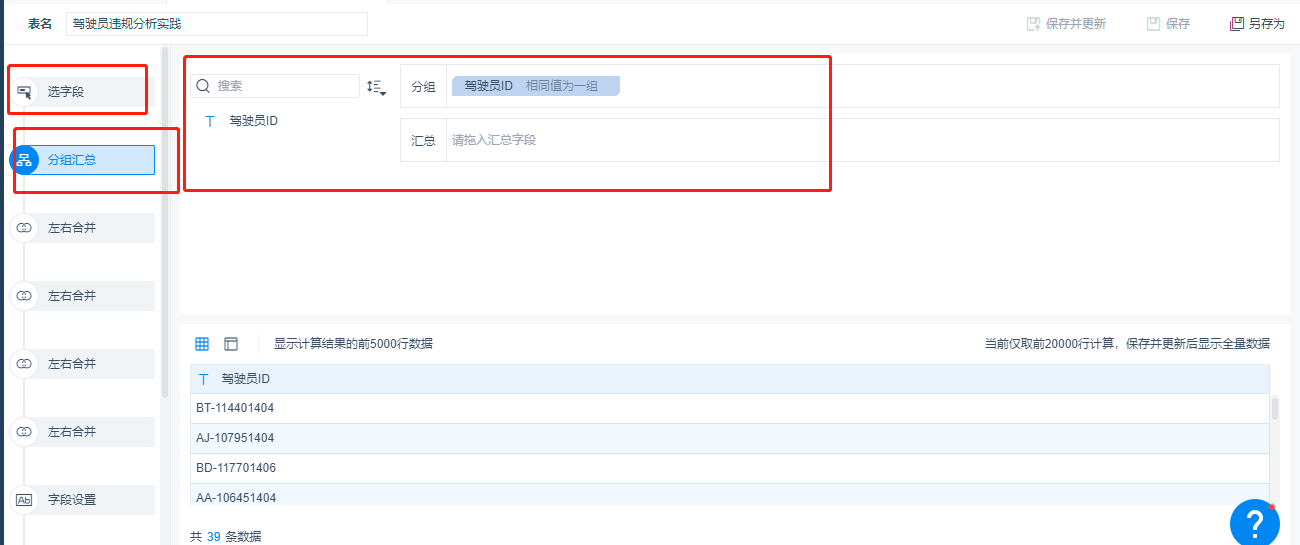
4）驾驶员第二次违规时间：在去掉了每个驾驶员第一次违规的数据后，与得到最早一次违规时间的方法相同，即可得到



5）驾驶员违规分析实践

①选字段：选取驾驶员违规信息数据表的“驾驶员ID”字段

②分组汇总：去掉驾驶员ID的重复数据



③与“驾驶员最早一次违规时间”利用驾驶员ID进行左合并得到每个驾驶员的第一次违规时间

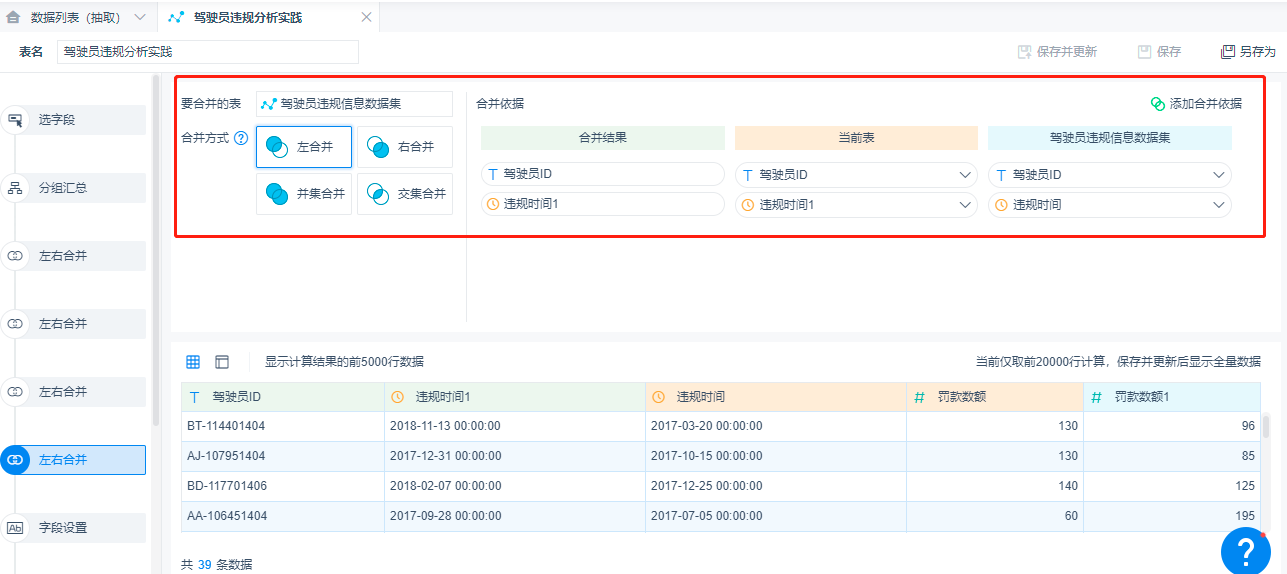


④与“驾驶员违规信息数据表”进行左合并，利用驾驶员ID和违规时间得到“第一次违规的罚款金额”



⑤同上，分别于“驾驶员第二次违规时间”与“驾驶员违规信息数据表”做左合并，得到第二次违规时间和第二次违规罚款金额

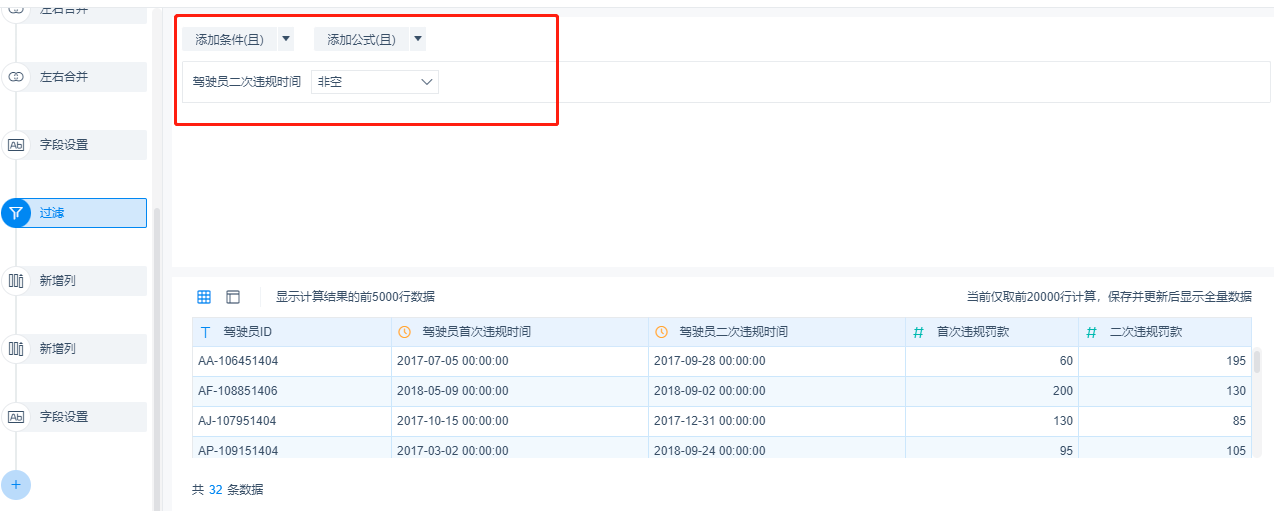




⑥字段设置，将字段依次命名为：驾驶员首次违规时间、驾驶员二次违规时间、首次违规罚款、二次违规罚款



⑦过滤：将第二次违规时间为空的过滤掉，不参与分析



⑧新增列：距首次违规间隔天数



⑨新增列：前两次罚款差额



三、最终结果

