2018拼多多(内推)数据分析师试题

## 问答题1

现有一个数据库表Tourists，记录了某个景点7月份每天来访游客的数量如下：

id date visits

1 2017-07-01 100

……

非常巧，id字段刚好等于日期里面的几号。现在请筛选出连续三天都有大于100天的日期。

上面例子的输出为：

date

2017-07-01

……

## 问答题2

在一张工资表salary里面，发现2017-07这个月的性别字段男m和女f写反了，请用一个Updae语句修复数据

例如表格数据是：

id name gender salary month

1 A m 1000 2017-06

2 B f 1010 2017-06

## 问答题3 （哪位大神教教我怎么做）

现有A表，有21个列，第一列id，剩余列为特征字段，列名从d1-d20，共10W条数据！

另外一个表B称为模式表，和A表结构一样，共5W条数据

请找到A表中的特征符合B表中模式的数据，并记录下相对应的id

有两种情况满足要求：

1每个特征列都完全匹配的情况下。

2最多有一个特征列不匹配，其他19个特征列都完全匹配，但哪个列不匹配未知

## 问答题4

我们把用户对商品的评分用稀疏向量表示，保存在数据库表t里面：

t的字段有：uid，goods\_id，star

uid是用户id；goodsid是商品id；star是用户对该商品的评分，值为1-5。

现在我们想要计算向量两两之间的内积，内积在这里的语义为：对于两个不同的用户，如果他们都对同样的一批商品打了分，那么对于这里面的每个人的分数乘起来，并对这些乘积求和。

例子，数据库表里有以下的数据：

U0 g0 2

U0 g1 4

U1 g0 3

U1 g1 1

计算后的结果为：

U0 U1 2\*3+4\*1=10

……

编写SQL做上述计算。

## 问答题5

微信取消关注分析，题目太长了，没记录。。