

中金财富证券互联网金融部笔试题目（两题都做）

一、SQL

将给到原始数据入库（已经给到建表 sql 语句，只要执行 sql 即可建表和插入数据），共三张表：1、页面访问数据：page_visit_test_201909。2、点击事件触发数据：event_visit_test_201909。3、订单数据：order_test_201909。

题目：写 sql 语句，统计：每个客户，在签约订单前（不包含签约的当日），在每个页面的访问次数，在每个点击事件的触发次数。

提示和说明：

- 1) 所统计的是：客户在签约（购买）投顾服务订单之前所访问的各个页面的次数+在客户在签约（购买）投顾服务订单之前所点击的各个事件的次数。
- 2) 某人的页面访问次数的统计：在页面访问表（page_visit_test_201909）中的任意一条数据，代表某客户（“客户编号”字段）在某日（“页面访问的时间”字段）访问了某页面（“访问的页面”字段）1 次，计数为 1。
- 3) 某人点击事件次数的统计：在点击事件表（event_visit_test_201909）中的任意一条数据，代表某客户（“客户编号”字段区分）在某日（“事件发生时间”字段区分）点击了某事件（“事件名称”字段区分）1 次，计数为 1。
- 4) 最终展示将每个客户访问各页面的次数和点击各事件的次数同时展示在一个结果表中，展示形式如下图所示：

Message	Result 1	Profile	Status								
客户编号	找投顾	选组合	投顾观点详情页	投顾观点评论页	投顾观点列表	投顾个人页	投顾组合评论页	投顾组合详情页	互联网投顾首页	是否购买	
▶ 30247447	1	1	1	0	0	1	0	0	0	1	
10177017	2	1	2	0	0	0	0	0	0	1	

- 5) 上图中最后一个字段“是否购买”的判断条件是：如果在订单数据中能够找到对应的客户（编号），且“组合名称”不为空，则代表该客户签约（购买）了投顾服务（产品）。
- 6) 如果遇到重复购买的人，则第一次购买前的行为数据为：9 月 1 日至第一次购买前的行为数据（统计）；第二次购买前的行为数据为：第一次购买和第二次购买之间用户行为数据（统计）；以此类推。
- 7) 上述 sql 完成后，需在 mysql 中能直接运行。
- 8) 提交的文档中除 sql 代码外，将跑出的结果截图附上。

二、Python

读取数据源（adult），并且在程序中输出如下信息：

- 1、一共多少条数据记录？程序中统计并输出，输出格式：“数据总数：XX 条”。
- 2、每个字段中的异常值（含缺失值）分别有多少？程序中统计并输出。
输出格式：XX 字段：空值 XX 个，aa 异常值 YY 个，…，异常值总计 ZZ 个；YY 字段：空值 XX 个，bb 异常值 YY 个，…，异常值总计 HH 个；
- 3、对异常数据进行处理，然后将数据分为建模集（70%）和预测集（30%）（说明：自行编写一个样本划分的方法，不可调用现有的包）；
- 4、Label 字段为因变量（Y），从其他字段中挑选合适的变量做为自变量（X），建立预测模型，并使用预测集进行模型验证。输出模型的评价指标（预测准确率（precision）、召回率（recall）、AUC 值），并绘制 AUC 曲线（图像）。（建模与评价可直接调包实现）

5、将上述 1-4 所有的结果，使用程序统一输入到一个 Word 文档中。

6、提交题目的时候，既要提交 word 文档，也要提交源代码 (Python)。