

百川智能-2024 校园招聘笔试题目

题目说明

1. 不限 SQL 类型，常见数据库 SQL 均可(MySQL、Hive、SparkSQL、Clickhouse、StarRocks...)
2. 每题可简要说明实现逻辑
3. 可有多种实现方式（可选）

笔试题 1

很久很久以前，有一个 NBA 球探，主要负责在大学联盟中，给球队寻找有潜力的年轻射手。现在他手里有非常多而且格式凌乱的新秀投篮数据，让球探不知如何选择，作为球探的助理，你需要给出一份符合要求的球员名单。要求如下：

1. 在一个学期的大学比赛中，至少连续登场 5 次，并且每场比赛的前 5 次投篮的总分是连续增长的
2. 在一场比赛中既投中过 2 分球也投中过 3 分球

数据表名称为 player_score，表内容示例如下(每周有且只有一场比赛)：

player_name	score	score_ts
player_a	2	2023-09-10 10:03:10
player_b	2	2023-09-11 10:04:23
player_a	0	2023-09-10 10:04:23
player_a	2	2023-09-10 10:03:53
player_b	3	2023-09-13 10:04:22
player_c	0	2023-09-10 10:04:23
.....		

字段含义：player_name: 球员名称、score: 每次投篮得分、score_ts: 每次投篮时间

笔试题 2

用户访问 App 日志表(user_log)，表中包含用户 ID、登陆时间、退出时间。表内容示例如下：

uid	login_ts	logout_ts
1	2023-09-10 00:03:10	2023-09-10 00:13:15
2	2023-09-10 11:12:56	2023-09-10 12:19:22
1	2023-09-10 23:03:10	2023-09-10 23:45:07
.....		

字段含义：uid: 用户 ID、login_ts: 登陆时间、logout_ts: 退出时间

问题 1:

求每一天，最多有多少人同时使用 App

问题 2:

计算 App，近半年内，每一天的第 7 日留存率

tips1: 第 7 日留存率是指用户在第 1 天登录后，第 7 天也登录了 App 的用户比例。例如，如果有 100 个用户在第一天登录，其中有 80 个用户在第 7 天仍然登录，那么 7 日留存率就是 $80 / 100 = 0.8$ ，也就是 80%。

笔试题 3

在大语言模型评测场景中、为了评价两个模型的效果，通常会让两个模型打比赛。比赛结果计分方式为：胜出得 3 分，打平得 1，负得 0 分。实际测评时，通常会使用多个数据集并且在打多轮比赛之后，综合对比来评估模型效果。模型比赛的结果保存在表 evaluation 中。示例如下：

表 evaluation 第一条记录含义，模型 algo_model1_v1 和模型 algo_model1_v2，使用 data_set_v1 数据集，打了一轮比赛，"DIM1"维度上模型 algo_model1_v2 胜出， "DIM2"维度上平局，"DIM3"维度上模型 algo_model1_v1 胜出

data_set	model1	model2	DIM1	DIM2	DIM3
data_set_v1	algo_model1_v1	algo_model1_v2	负	平	胜
data_set_v2	algo_model1_v1	algo_model1_v2	胜	负	负
data_set_v1	algo_model1_v1	algo_model1_v3	负	胜	平
data_set_v3	algo_model1_v2	algo_model1_v3	平	负	平
.....					

字段含义

data_set: 数据集、model1：模型 1、model2: 模型 2、DIM1: "DIM1"维度比赛结果、DIM2: "DIM2"维度比赛结果、DIM3: "DIM3"维度比赛结果

问题：请计算每个数据集下、每个维度下，得分最高的三个模型（只保留这三个模型），且要求排序：数据集排序 v2->v3->v1，维度排序 DIM3->DIM1->DIM2，按得分降序

结果示例:

数据集	维度	模型名称	总比赛 场数	平均分	胜场数	负场数	平场数
data_set_v1	DIM1	algo_model1_v1	36	1.83	21	12	3
data_set_v1	DIM1	algo_model1_v2	6	3	6	0	0
data_set_v1	DIM2	algo_model1_v1	11	1.63	6	5	0
data_set_v1	DIM3	algo_model1_v1	6	2.5	5	1	0
data_set_v2	DIM1	algo_model1_v1	7	1.71	4	3	0
data_set_v2	DIM2	algo_model1_v1	4	3.25	3	0	1
.....							