SQL Database

1. Friend Requests I: Overall Acceptance Rate

Table: `friend_request`

sender_id	send_to_id	request_date
1	2	2016_06-01
1	3	2016_06-01
1	4	2016_06-01
2	3	2016_06-02
3	4	2016-06-09

Table: `request_accepted`

requester_id	accepter_id	accept_date
1	2	2016_06-03
1	3	2016-06-08
2	3	2016-06-08
3	4	2016-06-09
3	4	2016-06-10

Write a query to find the overall acceptance rate of requests rounded to 2 decimals, which is the number of acceptance divided by the number of requests.

For the sample data above, your query should return the following result.

Note:

- The accepted requests are not necessarily from the table friend_request. In this case, you just need to simply count the total accepted requests (no matter whether they are in the original requests), and divide it by the number of requests to get the acceptance rate.
- It is possible that a sender sends multiple requests to the same receiver, and a request could be accepted more than once. In this case, the 'duplicated' requests or acceptances are only counted once.
- If there are no requests at all, you should return 0.00 as the accept_rate.

2. Friend Requests II: Who Has the Most Friends

requester_id	accepter_id	accept_date
1	2	2016_06-03
1	3	2016-06-08
2	3	2016-06-08
3	4	2016-06-09

Write a query to find the person who has most friends and the most friend number. For the sample data above, the result is:

id	num
3	3

Note:

- *It is guaranteed there is only 1 people having the most friends.*
- *The friend request could only been accepted once, which mean there is no multiple records with the same requester_id and accepter_id value.*

Follow-up: In the real world, multiple people could have the same most number of friends, can you find all these people in this case?

3. Second Highest Salary

Write a SQL query to get the second highest salary from the `Employee` table.

```
+----+--------+
| Id | Salary |
+----+--------+
| 1  | 100    |
| 2  | 200    |
| 3  | 300    |
+----+--------+
```

For example, given the above Employee table, the query should return `200` as the second highest salary. If there is no second highest salary, then the query should return `null`.

```
+---------------------+
| SecondHighestSalary |
+---------------------+
| 200                 |
+---------------------+
```

Answer:

4. Department Top Three Salaries

The Employee table holds all employees. Every employee has an Id, and there is also a column for the department Id.

```
+----+-------+--------+--------------+
| Id | Name  | Salary | DepartmentId |
+----+-------+--------+--------------+
| 1  | Joe   | 70000  | 1            |
| 2  | Henry | 80000  | 2            |
| 3  | Sam   | 60000  | 2            |
| 4  | Max   | 90000  | 1            |
| 5  | Janet | 69000  | 1            |
| 6  | Randy | 85000  | 1            |
+----+-------+--------+--------------+
```

The Department table holds all departments of the company.

```
+----+----------+
| Id | Name     |
+----+----------+
| 1  | IT       |
| 2  | Sales    |
+----+----------+
```

Write a SQL query to find employees who earn the top three salaries in each of the department.

5. Customers Who Never Order

Suppose that a website contains two tables, the Customers table and the Orders table. Write a SQL query to find all customers who never order anything.

Table: `Customers`.

```
+----+-------+
| Id | Name  |
+----+-------+
| 1  | Joe   |
| 2  | Henry |
| 3  | Sam   |
| 4  | Max   |
+----+-------+
```

Table: `Orders`.

```
+----+------------+
| Id | CustomerId |
+----+------------+
| 1  | 3          |
| 2  | 1          |
+----+------------+
```

6. |user_id|question_id|question_order||action|timestamp|

Find the question id with highest answer rate:

```
Input:
+------+-----------+--------------+------------+-----------+------------+
| uid  | action    | question_id  | answer_id  | q_num     | timestamp  |
+------+-----------+--------------+------------+-----------+------------+
| 5    | show      | 285          | null       | 1         | 123        |
| 5    | answer    | 285          | 124124     | 1         | 124        |
| 5    | show      | 369          | null       | 2         | 125        |
| 5    | skip      | 369          | null       | 2         | 126        |
+------+-----------+--------------+------------+-----------+------------+
Output:
+-------------+
| survey_log  |
+-------------+
|    285      |
+-------------+
Explanation:
question 285 has answer rate 1/1, while question 369 has 0/1 answer rate, so output 285.
```

Follow up:

- How to dynamically change the order of the questions showing to the users to achieve the highest conversion rate
- What should we do to the questions with only a few ‘show’ records.

7. **Advertiser spend and ROI metrics**

Given the following two tables:

- Advertiser info table with columns `advertiser_id`, `ad_id` and `spend', where` spend` is what the advertiser paid for this ad.
- Ad info table with columns `ad_id`, `user_id` and `price`, where `price` is how much the `user_id` spend through this ad., assuming all prices > 0.

Questions

- The fraction of advertisers has at least one conversion.
- What metrics would you show to advertisers.

Follow up question:
In which case ROI is not the best metric?
Answer: when advertiser cares less about revenue but more about CTR (eg. marketing campaign), ROI is not the best metric

8. **Create daily tracking table of user status**

Given a table that each day shows who was active in the system and a table that tracks ongoing user status, write a procedure that will take each day's active table and pass it into the ongoing daily tracking table. Possible states are:

- user stayed (yesterday yes, today yes)
- user churned (yesterday yes, today no)
- user revived (yesterday no, today yes)
- user new (yesterday null, today yes)

Note: you'll want to spot and account for the undefined state.

TABLE1: Tracking {user, status}
TABLE2: Day {user}

How do you calculate monthly active users, churned users and resurrected users from a user activity log with userID and DateTime

9. **TABLE Friend: {datetime | action: ('request_sent', 'request_accepted') | actor_id | target_id}**

*Find friend acceptance rate **trending***

- Define how long you have to wait before a friend request is considered rejected (e.g. 1 week)
- Here a user may send multiple request to a same user at different time

10. **Recommend pages your friends liked.**

You have two tables

- the first table has data about the users and their friends.
- the second table has data about the users and the pages they have liked.

TABLE1: Friends {user_id, friend_id}
TABLE2: Page {user_id, page_id}
Write an SQL query to make recommendations using pages that your friends liked. The query result should not recommend the pages that have already been liked by a user.

11. Advertise Conversion Rate

TABLE1: Advertise {ads_id, timestamp, publisher_id, price}

TABLE2: VIEW {ads_id, timestamp, view, clicks}

QUESTION: conversion rate for publisher_id = 'P' and Date = 'date' (a specific publisher on a specific date)

12. table: date | user_A | user_B | number_messages

Each row has the number of messages between a unique user pair

1) *what can you tell from this table*
2) *Write a query for the distribution of number of messages for each user. How the distribution is gonna look like and why?*
3) *Write a query find the top partner for each user (most messages)*
4) *SUM(n_msg_with_top_partner)/SUM(all_messages_with_all_contacts），sum over all users*

13. Table name: content_actions

{user_id | content_id | content_type ('post', 'photo', 'comment') | target_id}

#story: post or photo

1) *Generate a distribution for the #comments per story.*
2) Does this count for stories with 0 comments?

14. TABLE: {time, user_id, app_id, event ('imp' or 'click')}.

Write a query for CTR.

15. FB -(send verification code)-> mobile carriers -(send as sms)-> user -(confirm)-> FB

TABLE1:sms_send

{ds(date)|carrier|country|phone_num|event_type}

event_type: confirm/recovery, etc...

TABLE2: confirm

{ds|phone}

- How many requests fb sent to each carrier yesterday? (sql)
- If the confirmation rate decreased by x%, what might be the reason?

- Assume the number of messages FB sent don't change, but confirmation rate decreased by x%, why?
- Assume carrier is the reason for confirmation decrease, how to find which carrier? (sql)
- If multiple carrier are down at the same time, FB needs to contact 5, how to choose?

16. Table {post|comment} find the distribution of comment

17. Comment on post

user_data

user_id	current_link_id	previous_link_id	data_type
1	2131		post
2	53123		picture
3	213		movie
6	1231	2131	comment
11		53123	comment
12		53123	comment
13		53123	comment
14		53123	comment

1) *percent of post having at least one comment*

2) *percent of post having at least 5 comments*

18. TABLE {content_id, content_type (comment/post),target_id}

- If content_type='comment', then target_id is content_id(post)
- If content_type='post', then target_id is NULL

1) *What is the distribution of the comment?*

2) *If now content_type becomes (post, video, photo, article), calculate the comment distribution of each content_type*

Mock Data and Questions:

This example doesn't pertain to Facebook but is representative of the data and questions you may see:

An attendance log for every student in a school district
attendance_events : date | student_id | attendance
A summary table with demographics for each student in the district
all_students : student_id | school_id | grade_level | date_of_birth | hometown

Using this data, you could answer questions like the following:

- What percent of students attend school on their birthday?
- Which grade level had the largest drop in attendance between yesterday and today?

Simplest SQL query to find the second largest integer value in a specific column

```
SELECT MAX( col )
 FROM table
WHERE col < ( SELECT MAX( col )
              FROM table )
```

1. 有一张表，有 time, user_id, app, event (impression，click) , 每个用户在每个 app 上有一定几率弹出一个窗口填写信息，如果填写了 event 为 click，如果只看见没填写为 impression，没看见为空。求这个功能的 click through rate。加问如果 CTR>100%是什么原因？如果现在每一个 impression 可能对应多个 click，如何从所有 click 记录里选出正确的那个记录来计算？
2. Survey: saw/skip/answered. Along with question_order & question_id
   If a new user skips the question with highest & second highest frequently answered question, how should we change the order of questions so we can get more answered question (highest conversion rate)?
3. advertiserid, adid, spent, 另外一个表格 uid, adid, purchase， 要求过去三十天 advertiser 花了多少钱在 ad 上 per advertiser，distribution 是什么，why？ 然后 by adid 求 ROI
4. Post, comment 那道题，地里出现了无数次了，之前看到过，还比较顺手。不难。
   1)求 comment distribution
   2)求 comment distribution by content type
   3)如果不看 date range, data 太大怎么办。我说就看今天的，他问那今天的有什么问题。就是没法 capture cumulative num of comments, 只有今天的。
   4) 你现在有# of comments for a certain post, 你怎么知道这个 number 是 reasonable 的。--取些 sample 看 variance, confidence interval
5. week over week change％
6. friending trend, action_id | target_id | action{‘sent’,’accept’,’unfriend’} | date
   要求计算每个人的好友. 如何判断两个人是不是好朋友
7. Are you sure you want to buy <item> from <game>?
   [YES][NO]
   table dialoglog
   timestamp, userid, game, event {'impression', 'yes'}
   what is the CRT， what if there is a lot of pollution in this table

8. There is a table that tracks every time a user turns a feature on or off, with columns user_id, action ("on" or "off), date, and time.
   1) How many users turned the feature on today?
   2) How many users have never turned the feature on?
   3) In a table that tracks the status of every user every day, how would you add today's data to it?

9. Write a sql query to find out the overall friend acceptance rate for a given date?
   Table :- User_id_who_sent|User_id_to_whom|date|Action (Sent, accepted, rejected etc)