Competition #1: Data Analytics

Multi-Channel Contacts Problem

6 March 2021

Competition Details

Start time: 6 March 2021, 1pm (GMT+7) / 2pm (GMT+8)

End time: 6 March 2021, 4pm (GMT+7) / 5pm (GMT+8)

Duration: 3 hours

Table of Contents

Background

Customer service is an important element of the Shopee business, as providing a good service for our customers end-to-end is critical for business growth and brand image. Our goal is to resolve the customer's issue within the least amount of time while requiring the least amount of customer effort.

One measure for customer effort is the number of times a customer has to approach customer service over a particular issue, this is also known as the metric "Repeat Contact Rate" or RCR. Shopee is interested in studying the RCR in order to improve the effectiveness of our customer service.

Customers can contact customer service via various channels such as the livechat function, filling up certain forms or calling in for help. Each time a customer contacts us with a new contact method, a new ticket is automatically generated. A complication arises when the same customer contacts us using different phone numbers or email addresses resulting in multiple tickets for the same issue. Hence, our challenge here is to identify how to merge relevant tickets together to create a complete picture of the customer issue and ultimately determine the RCR.

Task

For each ticket, identify all contacts from each user if they have the same contact information.

For the purpose of this question, assume that all contacts from the same Phone Number / Email are the same user.

Basic Concepts

Each Order ID represents a transaction in Shopee.

Each Id represents the Ticket Id made to Shopee Customer Service.

All Phone Numbers are stored without the country code and the country code can be ignored.

Contacts represent the number of times a user reached out to us in that particular ticket (Email, Call, Livechat etc.)

If a value is NA means that the system or agent has no record of that value.

Example

Ticket A

Id	0
Email	John@gmail.com
Phone	NA
Order ID	12345678
Contacts	5

Ticket B

Id	1
Email	NA
Phone	682212345
Order ID	12345678
Contacts	2

Ticket C

Id	34567
Email	Wick@gmail.com
Phone	682212345
Order ID	NA
Contacts	4

Ticket D

Id	78999
Email	Wick@gmail.com
Phone	NA
Order ID	NA
Contacts	3

Explanation

Each of these tickets are related either directly or indirectly through Email, Phone or Order ID, therefore each ticket belongs to the same user.

- Ticket A and B are linked through Order ID
- Tickets B and C are linked through Phone
- Tickets C and D are linked through Email
- Tickets A and D are indirectly linked through tickets A > B > C > D

In this example, this user has a total of 14 contacts. The ticket_trace/contact pair for this user would be 0-1-34567-78999, 14.

Submission Format

Determine how many contacts each user has had across the various tickets.

Two columns required:

- ticket_id
- ticket_trace/contact
    - Format: <tickets in ascending order>, <total contact>

For each ticket, identify all other ticket_ids that belong to the same user, **sorted in ascending order**, as well as the total contacts the user had. If two tickets belong to the same user, the value for uid/contact for both tickets should be identical.

ticket_id	ticket_trace/contact
0	0-1-34567-78999, 14
1	0-1-34567-78999, 14
2	2-2456-4323-5343-6789, 13

Your submission should have **500,000 rows** (excluding headers), each with 2 columns.