Jerry.ai Data Analyst Question

Haomin Lin

1 Assumptions

In order to analyze the two tables and generate insights about whether the current plan is working well or not, I need to make following assumptions:

1. In 'User' Table, I assume the 'referring_user_id' corresponds to another user that referred the current user to Jerry.ai. This means the 'referring_user_id' can be NULL.
2. Because the date when a user is created is not included, I assume the day they pay the bill as the day their accounts are created.

2 Step 1: Analysis

In order to derive metrics from data to help decide whether this program is running well or not, I plan to run information extraction as:

2.1 Size of Referred Users by Month

Firstly, we want to know how the size of new referred users varies compared with number of new users by month. From this analysis, we can see if the increase size of referred users is attributed to the increase of referring users or the attraction of the reward program. The SQL query is:

```
SELECT num_referee/num_referer as refer_rate, year_month
FROM
((SELECT count(DISTINCT U.id) as num_referee, FORMAT(date,'y') as year_month
FROM User U JOIN Purchase P
on U.id = P.user_id
WHERE U.referring_user_id IS NOT NULL
GROUP BY FORMAT(date,'y')) AS referred
join
(SELECT count(DISTINCT U.id) as num_referer, FORMAT(date,'y') as year_month
FROM User U JOIN Purchase P
on U.id = P.user_id
GROUP BY FORMAT(date,'y')) AS referring
ON referred.year_month=referring.year_month) AS refer;
```

refer_rate can get a picture of how attractive user referral program is on a monthly basis.

2.2 Amount of Money Brought by Referral

In order to evaluate the effect of our referral program, we need to analyze whether the users brought in by referral are positive about spending money at Jerry.ai. So we need to observe how much money they spend on a monthly basis. The SQL query is:

```
SELECT SUM(total-discounts) as referee_payment, FORMAT(date,'y') as year_month
FROM Purchase
WHERE user_id IN
(SELECT id
FROM User
WHERE referring_user_id IS NOT NULL)
GROUP BY FORMAT(date,'y');
```

the referee_payment values can show us how the money users paid by referred users varies month to month.

3 Step 2: Recommendation

Given different analysis results from data, the program should be continued give the following cases:

1. When refer_rate is increasing. In this case, it means our brand has been built among users since we offer the same benefit but attract more people on the condition that same number of users offer referral. In this case, as long as the referee_payment is not too low, keep building our brand is beneficial.

2. When refer_rate is stable but at a high level, even if it's not increasing, high refer_rate means this program is still productive, and keep running it can keep bringing in more customers.

3. When refer_rate is decreasing from a high level, but referee_payment is still over a threshold. As long as this is not damaging, I think keep the possibility of attracting new users is still a good thing, and this is also when we explore new strategies before the refer program loses all its power.

In the other cases, the program needs to be discontinued. Specifically:

1. When refer_rate is stable or decreasing at a low level. This case is just equal to giving out money, as we cannot attract enough users to support our program, it's time to switch to another plan.

2. When referee_payment is less than a threshold. When this happens, it means even if our program can bring in a lot of new referred users, they actually don't contribute to our project, and won't help us gain more payment. This observation would suggest us try other methods to make new users willing to pay for our project.