Question Type: Product

Duration: 40 Minutes
Difficulty: Medium
Domains: Product

Problem

Address the following questions for the Netflix data scientist position.

#1 - How would you improve conversions on Netflix?

#2 - How would you test your hypothesis?

Solution

> **#1 -** How would you improve subscriptions on Netflix?

[Candidate] Before I propose a solution, I would like to understand the conversion funnel on Netflix. My understanding is that new users are funneled into subscriptions in three steps:

1. Site Visit - New users enter the Netflix landing page to learn about features.
2. Sign-Up - Users create accounts with email addresses and payment details.
3. Subscribe - Users gain access to movies and shows in monthly subscriptions.

[Interviewer] You have the right idea about the Netflix sign-up flow. What opportunity do you think can improve subscriptions?

[Candidate] I would imagine that each of the three steps poses an opportunity to improve subscriptions in terms of acquisition and retention. Making the subscription more appealing in the landing site and reducing friction on the sign-up flow can improve the conversion rate (acquisition). Adding more relevant contents and improving recommendations can prevent user churns (retention).

[Interviewer] Seems like you have a comprehensive list. Can you choose one aspect to improve?

[Candidate] I believe there's an opportunity to improve the sign-up flow, thereby increasing the conversion rate. Can I assume that the current sign-up process requires credit card information?

[Interviewer] Yes.

[Candidate] It's my understanding that the sign-up process should remove as much friction as possible unless the users are bad actors. Requiring users to pay first before content access may provoke some users to rescind the sign-up.

[Interviewer] That's an interesting insight. What would be the alternative then?

[Candidate] Instead of requesting payments first, perhaps users should enroll in a trial period - 1-month free. If they want the access beyond the trial period, payments are requested then.

[Interviewer] What segment of new users would benefit from this free trial? Also, do you forsee any risks?

[Candidate] When it comes to user acquisitions, users can be segmented based on their willingness to convert. High intent users will usually convert regardless of friction. For instance, they entered the Netflix site with intentions of signing-up. Low intent users require persuasion and prone to exiting the conversion funnel upon frictions. The free trial would benefit the low intent users as they can sample the product before committing to purchase.

On your second question, I do see a potential risk. Bad actors may abuse the free trials, cancelling the membership before the subscriptions and re-enrolling with new accounts to continue using the service for free.

Interviewer Solution

This is a question that tests the candidate's product-sense of Netflix product. Understanding the lifecycle of users on a subscription-based product - (1) Awareness (2) Acquisition (3) Retention - are fundamental to address this question.

Note that this style of problem can be plugged into any other tech company interviews. For instance, on LinkedIn, you can imagine the following question: "How would you improve LinkedIn's premium subscription?"

With that basis, the candidate must define *what to improve* in the subscription flow - does it involve personalized email campaigns on subscription features (awareness), reduced friction on the subscription process (acquisition), or adding features to improve engagements (retention)?

A candidate should choose one product, think about the lifecycle of a user and brainstorm a way to improve it.

In this example above, the candidate suggested including a trial period before inputing a credit card to incentivize low-intent users to subscribe. Note that the candidate was very specific on what to improve, how to improve it and whom to target:

- **What:** Sign-up flow
- **How:** Free Trial
- **Whom:** Users with low intent

This was a decent idea. However, the response could use some improvements:

First, the candidate should clearly define the metric(s). When using the "conversion rate," what's the numerator and denominator of the formula?

Depending on the context, the conversion rate could be the following:

1. # of subscribers / # of new users in the sign-up flow

2. # of subscribers / # of new visitors

Another suggestion is the duration of the trial. The candidate proposed 30-days for the free-trial duration. This means that the entire experiment will take 1 to 2 weeks to gather users + a 30-day window to gather the experimental result. Generally, the business team wants to make decisions quicker. Instead of 30-day free trial, perhaps it's ideal to start with a 7-day trial to run a quicker experimentation.

> **#2 -** How would you test your hypothesis?

[Candidate] Since this is a comparison of two groups - control (no trial) vs treatment (30-day trial) - and the metric is the conversion rate, the statistical test is a T-test for two proportions.

[Interviewer] What's the formula of your conversion rate?

[Candidate] The conversion rate is the number of users who enrolled in the monthly subscription over the new visitors.

[Interviewer] How would you determine your sample size?

[Candidate] I would estimate the sample size based on the significance level. A reasonable significance level in AB testing is 0.05.

[Interviewer] How long would you run your experimentation?

[Candidate] It will be based on the sample size per group and traffic allocation to the experiment. For instance, suppose that a sample size of each group 10,000 users, totaling 20,000 users required for the experiment. Based on the user traffic, 5,000 users can be allocated to the experiment per week. Then, the experimentation time will be 4 weeks.

[Interviewer] Will four weeks be a reasonable for testing your hypothesis?

[Candidate] Yes.

[Interviewer] How would you interpret your result?

[Candidate] If the p-value is less than the significance level, reject the null hypothesis that the sign-up has no effect on conversions. Otherwise, fail to reject and conclude that there is no effect.

[Interviewer] Suppose that you failed to reject the null hypothesis, what's next?

[Candidate] Then the idea on credit card fails and move onto another idea on improving conversions.

Interviewer Solution

This question asks the candidate to test the idea provided in the previous problem.

A comprehensive solution should detail the following:

1. Hypothesis Testing - *What feature is tested? What's the hypothesis? What's the control and treatment(s)? Which population of users are being tested (i.e. all users, new visitors, or users in sign-up flow)? What's the metric? What's the statistical test?*
2. Experiment Design - *What's the experimental unit (i.e. users, device)? What's the significance level and power? What's the minimum detectable effect? How's the randomization performed? What's the sample size? How long should the experiment run? How will it handle extraneous factors such as seasonality, novelty and network effects?*
3. Result Interpretation - *What checks should you run to ensure that the experiment ran error-free? What's the statistical outcome?*
4. Decision Rule - *Based on statistical outcome and practical significance, should the business launch the new feature? What's the cost-benefit analysis of launching the new feature?*

Overall, the candidate's responses were underwhelming with flawed ideas and superficiality.

First, the candidate rushed to statistical testing without defining the pre-requisites (i.e. metric, user segment, hypothesis statements and e.t.c.).

"Since this is a comparison of two groups - control (no trial) vs treatment (30-day trial) - and the metric is the conversion rate, the statistical test is a T-test for two proportions."

Also, the candidate's choice of metric is flawed as defined the conversion rate as "the number of users who enrolled over the number of visitors." The better choice for the denominator would be the number users in the sign-up flow. Otherwise, the metric will be quite noisy (i.e. new visitors who visit the landing site but doesn't enter the sign-up flow).

"I would estimate the sample size based on the significance level. A reasonable significance level in AB testing is 0.05. "

The candidate doesn't seem to have a clear understanding of the sample-size determination. A sample size is determined based on the significance level, power, and MDE. The candidate only mentioned the significance level. In general, the significance level should be 0.05, power = 0.80 and MDE = 1% for a platform with a large user base like Netflix.

"...the experimentation time will be 4 weeks."

This is too long for an experiment time. Ideally, the experimentation time should be 1 to 2 weeks. In this case, the candidate is proposing that the experimentation runs 4 weeks. One other element that the candidate forget to mention that to gather the result, it will be adidtional 4 weeks (~ 30 days) given that users are on the free trial period. In total, it would be 8 weeks! The candidate should increase traffic allocation to reduce the experimentation time to 1 to 2 weeks.

“If the p-value is less than the significance level, reject the null hypothesis that the sign-up has no effect on conversions. Otherwise, fail to reject and conclude that there is no effect.”

The candidate’s response lacks depth. I’ve tried guiding him toward details when I asked, “what’s next?” after the experiment fails to rejection. However, he failed to produce a proper response. The candidate should mention diagnostics to ensure that the experiment ran without errors, recommend whether the hypothesis test should be calibrated, or launch the new feature. For instance, if the test rejects, perhaps the candidate could propose an alternative treatment to evaluate - add credit card in the beginning of the sign-up flow, followed by the free-trial (ideally 7 days to start). Additionally, the candidate needs to discuss the trade-off of launching. Just because an effect of a new feature is statistically significant, doesn’t mean that it should be launched if the practical signifiance is lacking. Meaning, the lift generated from the new feature is lacking compared to the cost of launching it in terms of engineering, UI, and business model changes.

Interviewer Assessment

In the statistics section, a candidate is assessed based on correctness and soundness of statistical methodology, product sense and communication. For each dimension the candidate is rated in the following scale: (5) superior, (4) good, (3) adequate, (2) marginal, (1) not competent.

Assessments	Rating	Comments
Statistical Methodology	2	The candidate's response was inadequate with false statistical methodology and conclusions. For instance, when asked about the sample-size determination, the candidate made a novice mistake, only mentioning the significance level as the parameter for the sample-size calculation. The candidate also seems to lack a reasoneable experimentation time. A total of 8 weeks (4 weeks of user assignment to control or treatment + 4 weeks of trial period) is far too long when a product decision needs to be tested and determined quickly. Lastly, the candidate seems to lack the understanding of a proper interpretation of statistical result.
Product Sense	4	The candidate seems to understand the lifecycle of a user on Netflix. He had a decent idea about using a "test before pay" approach to incentivize low-intent users to subscribe. This suggests that the candidate has conducted some research the Netflix product before the interview. One weakness in the candidate's response is the vagueness on metric. Just saying increase "the conversion rate" is insufficient. The formula needs to be clearly defined.
Communication	3	Although the candidate provided decent responses with clarity in the first half of the interview. In the second half, the responses were often superficial. The candidate should try to be concise with his message and comprehensive with details.