

A/B Testing Project

Experiment Overview: Udacity Free Trial Screener

by Wentian Ma

Udacity courses have two options on the course overview page: "access course materials" and "start free trial". If the student clicks "access course materials" button, they can view the class materials and take the quizzes for free, but they will not receive coaching support or a verified certificate which would generate additional revenue for Udacity. If the student clicks "start free trial" button, they will be asked to enter their credit card information, and then they will be enrolled in a free trial for the paid version of the course.

This experiment will test a change to ask student how much time they will devote to the course after they click "start free trial" button. If the student will devote less than 5 hours per week, Udacity will send a message suggesting that the student might prefer access course materials for free. Then, the student will have the option to continue enroll the course or take the suggestion and take the course for free.

The ultimate goal for this experiment is to test whether the suggestion will reduce the number of student who becomes overwhelmed because of lack of time without significantly reducing the number of students complete the course. Thus, Udacity could improve the overall student study experience and enhance coaches' capacity to support the right students.

Experiment Design

Metric Choice

In this experiment, I chose the number of cookies, the number of clicks, and the click-through-probability as Invariant Metrics. Since the unit of diversion is cookie, thus I chose click-through-probability instead of click-through-rate which rate is better in usability of a function. In detail, the Click Through Rate is full of useless clicks and pageviews like incrementation due to refreshing the page, multiple clicks. While the Click Through Probability will get rid of all those useless clicks and views to keep only meaningful data.

As for Evaluation Metrics, I chose the retention rate (number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by number of user-ids to complete checkout) and Net Conversion Rate (number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button)

Measuring Standard Deviation

For each of evaluation metrics, the estimated standard deviation of the retention is 0.05495 and 0.01560 for the net conversion given the sample size of 5000 cookies visiting the course

overview page. In detail, based on the info sheet, there are 40,000 unique cookies to view the course overview page, not 5,000. And so I re-calculate the numbers to how they would look if it were 5,000 cookies viewing that page, assuming they stay proportional.

Number of enrollment = number of unique cookies * CTP * enrollment %

Standard Deviation = $\text{SQRT}(P*(1-P)/\text{Number of enrollment})$

Sizing

Number of Samples vs. Power

Next, by using the analytic estimates of variance, I calculate the total number of pageviews for both groups that I will need to power the experiment appropriately. Since this is a multiple metric A/B test and it is hard to assume two metrics are independent from each other, thus I will use the Bonferroni Correction to inflate type I error during my analysis phase. Therefore, the corrected significance level is 2.5%, which equals $5\% / 2$.

Duration vs. Exposure

Given the information about the total number of pageviews for both groups, I set the percentage of Udacity's traffic 7.6% and the length of experiment is 20 days.

Experiment Analysis

Sanity Checks

Before start the experiment, I start by checking whether my invariant metrics are equivalent between the two groups. If the invariant metric is a simple count, it should be randomly split between the 2 groups, which should be roughly 50% for both. Otherwise, I should construct a confidence interval for a difference in proportions, then check whether the difference between group values falls within that confidence level.

For each of my invariant metrics, give the 95% confidence interval, the observed value is in the confidence interval, which means the number of cookies and the number of clicks both pass the sanity check. Therefore, I could proceed to the rest of the analysis.

Result Analysis

Effect Size Tests

Next, for my evaluation metrics, I calculate a confidence interval for the difference between the experiment and control groups, and check whether each metric is statistically and/or practically significance. The the confidence interval of retention rate does not include 0, therefore, it is statistically significant. However the confidence interval does include the practical significance boundary, which means it is not practically significant. As for net conversion rate, it is neither statistically significant nor practically significant.

probability of pool = $(\text{number of experiment} + \text{number of control}) / (\text{total number of experiment} + \text{total number of control})$

d= probability of experiment - probability of control

standard of deviation = $\sqrt{\text{probability of pool} \cdot (1 - \text{probability of pool}) \cdot (1/\text{total number of control} + 1/\text{total number of experiment})}$

Sign Tests

For each of my evaluation metrics, I do a sign test using the day-by-day data, and p-value of both retention and net conversion rate are not statistically significant.

Recommendation

Among the two evaluation metrics, there is only one metric, which is the retention rate, is statistically significant. Both metrics are not practically significant. Also according to day to day analysis, there is no significant difference between control group and experimental group. Therefore, there is not enough evidence to reject the null hypothesis which means there is no obvious difference observed between control group and experience group. In general, the suggestion pop out after time devoted question will not reduce the number of student complete the courses.

Therefore, I will launch this project. In order to get better result, I will run a multiple test with small fraction of number in the same time. Besides, I will also make a survey to investigate user experience about this new feature. I may also run a follow-up experiment with more detailed information given.

