

Experiment Design

Metric Choice

Invariant Metrics and Evaluation metrics:

Number of Cookies: The number of cookies is the unit of diversion in this experiment. The experiment funnel goes from Page views to clicks to start free trial to enrollment and payment. The number of cookies from control to experiment does not change so the results from the conversion rates can be compared, which is why this is an invariant metric. It cannot be used as an evaluation metric since the experimental design is not expected to drive any changes in this metric from control to experiment groups.

Number of userids: This is not an invariant metric as we expect to see a change in the number of people who enroll by the design of the experiment, in that it reduces once the weekly hours of commitment is communicated clearly early on before the enrollment in free trial. This is not being used as an evaluation metric itself since the absolute value of user-ids simply indicates the number of enrolments that passed through the funnel. A relative value gives much more information about the effectiveness of the experiment, which is what we need to know here.

Number of clicks: The experiment where the user is asked about the number of available hours for study opens after the user clicks free trail. The events leading up to this are the same in control and experiment, so this metric is also an invariant. This cannot be used as an evaluation metric similar to number of cookies as the experiment does not cause any changes to it.

Click through probability: This measures the chance a user clicks on free trial, measured by the ratio of number of clicks and number of cookies as defined above. By design of the experiment, this metric is an invariant. This is not being used as an evaluation metric as we do not expect to see any changes in it in the experiment group.

Gross conversion: Is the ratio of number of enrollments in the free trial and the number of clicks. Since the number of enrollments is not an invariant metric, this also is not an invariant metric. We expect this number to reduce in the experiment group, therefore this is an evaluation metric

Retention: Is not an invariant since we expect the number of user-ids to complete check out and the number of user-ids to make a payment to vary from control to experiment. This can be used as an evaluation metric as we expect the ratio to change.

Net conversion: Is not an invariant metric since we expect the number of user-ids to stay enrolled past the 14 day boundary to change, but not by a significant amount, so we want to observe this metric, if the experiment achieved the desired results. This is an evaluation metric in this experiment as we want to see that there is not a significant reduction in the number of users who enroll past the free trial period.

Expectations from the evaluation metrics:

During the course of the experiment we expect gross conversion to decrease, the direction and the magnitude of the change will be looked at in the experiment. The net conversion is not expected to change ideally, but if there is a change we would like to see if it crosses the two significance boundaries in either direction. Retention metric is expected to increase as we hope to retain more

students by displaying the message. This change is expected to be significant enough both statistically and practically for the change to launch.

Variability

Gross conversion : 0.0202

Retention: 0.0549

Net conversion: 0.0156

Analytical estimates were used in this experiment but given time empirical estimates would be ideal to use for practical purposes. I don't expect the variability to change by a significant amount between empirical and analytical, since the underlying distribution is normal. The unit of diversion here is a cookie and the unit of analysis are the same, therefore it is unlikely that there will be a lot of variability between the two units hence the analytical estimate is good enough to use. In case of retention an empirical estimate would give better results than analytical estimate.

Sizing:

Number of samples vs power

I did not use the Bonferroni correction in the analysis. Since the evaluation metrics are not independent. Retention metric required a large number of pageviews which would take up a lot of time to run the experiment on Udacity so it is dropped from the evaluation metrics. The evaluation metric selected are gross conversion and Net conversion. For an alpha of 0.05 and beta of 0.2, the pageviews required are 685325.

Duration vs Exposure:

Fraction of traffic diverted to this experiment is 0.5 and the number of days the experiments needs to be run is 35. These numbers were arrived from the baseline statistics and the minimum number of pageviews required calculated earlier. The fraction of the traffic is arrived at such that the duration of the experiment is optimal enough.

Smaller and longer experiment run the opportunity risk that Udacity cannot run any other experiment during this time. The experiment does not change course content or load time so user experience is not affected by this experiment.

Experimental analysis

Sanity checks

Metric	Lower bound	Upper bound	Observed	Check
Number of cookies	0.4988	0.5012	0.5006	Passed
Number of clicks	0.4959	0.5041	0.5005	Passed
Click through prob	-0.0013	0.0013	0.0001	Passed

All the invariant metrics passed the sanity check.

Result analysis

Effect size tests

95% confidence interval for Gross conversion: -0.0291 to -0.0120. The difference is both statistically and practically significant.

95% confidence interval for Net conversion is -0.0116 to 0.0019. The difference was neither statistically significant nor practically significant.

Sign tests

Pvalue for Gross conversion – 0.0026. The sign test is statistically significant.

Pvalue for net conversion – 0.6776. The sign test was not statistically significant.

Summary:

Bonferroni correction was not used in this experiment as the metrics being measured are correlated. If gross conversion is high, net conversion is also likely to be high, this indicates a positive correlation. The alpha resulting from this method for this reason might be too conservative for practical purposes. Bonferroni correction is applied when using multiple metrics to reduce the chance of getting a false positive as the number of metrics increases. We need all the metrics, both gross conversion AND net conversion to satisfy our launch criteria, controlling for false positives here doesn't apply as it has minimal effect in changing the outcome of the experiment. Therefore Bonferroni correction is not applied here as it is an AND case

Effect size hypothesis tests yielded both a statistically and practically significant result for gross conversion. This indicates that displaying the message before starting free trial decreased the number of users that signed up free trial by a statically significant amount. The purpose of the experiment was to achieve this without significantly reducing the number of students to continue past the free trial to enroll in the courses. The metric that measures this is the net conversion which from the effect size test and sign tests did not show a statistically or practically significant change.

Recommendation

Based on the parameters and conditions set for this experiment, it can be concluded that introducing a message before signing up for free trial did reduce the number of people enrolling in the free trial without significantly reducing the number of who have stayed enrolled past the free trial period. Upon

taking a closer look at the net conversion effect size test, the lower boundary of the confidence interval is negative, which might indicate that this statistic has reduced in the experiment group, which is not a desirable outcome. Based on this observation my recommendation would be not launch and conduct a follow up experiment to test other ways to achieve the objective.

Follow up: Reduce early cancellations

The follow up experiment hypothesizes that the reduction in the number of students dropping out of free trial might be because they find the course material challenging because they do not have the prerequisite skills. The experiment here is to introduce a small quiz regarding the prerequisites for each course and the users who pass the prerequisite test can enroll in the free trial and the users who do not have a choice to be redirected to the right prerequisite course pages on Udacity or proceed with the free trial.

In this new experiment when the students clicks on the free trial a brief questionnaire about their current skills levels and prerequisites required for the course is presented to them. If the student does not meet the pre-requisites for the course a message suggesting if they would like to take the pre-requisite course pops up. At this point the student has the option to take the pre-requisite course or join the free trial, while indicating that the pre-requisite course is free and not included in the free trial period. The null hypothesis would be that the ratio of the number of students staying past the free trial period to the number of students who clicked on the free trial button does not change between control and experiment groups.

The user traffic funnel similar to the current experiment starts with the number of unique cookies to view the course page followed by number of click on the free trial button. Once the user enrolls they are tracked through user-id. Number of users who enroll in the free trial and number of users to remain past the free trial period. Gross conversion, retention and net conversion will be used as evaluation metrics. The invariant metric here would be the number of clicks on the free trial button

A successful experiment would be where we see an increase in gross conversion ratio, while the net conversion rates does not decrease by both statistically and practically significant levels, similar to the current experiment.