Project-7-AB-Testing

Udacity Design A/B Test

A/B Testing: Udacity Free Trial Screener

Experiment Overview: Free Trial Screener

At the time of this experiment, Udacity courses currently have two options on the home page: "start free trial", and "access course materials". If the student clicks "start free trial", 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. After 14 days, they will automatically be charged unless they cancel first. If the student clicks "access course materials", they will be able to view the videos and take the quizzes for free, but they will not receive coaching support or a verified certificate, and they will not submit their final project for feedback.

In the experiment, Udacity tested a change where if the student clicked "start free trial", they were asked how much time they had available to devote to the course. If the student indicated 5 or more hours per week, they would be taken through the checkout process as usual. If they indicated fewer than 5 hours per week, a message would appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free. At this point, the student would have the option to continue enrolling in the free trial, or access the course materials for free instead. This screenshot shows what the experiment looks like.

The hypothesis was that this might set clearer expectations for students upfront, thus reducing the number of frustrated students who left the free trial because they didn't have enough time—without significantly reducing the number of students to continue past the free trial and eventually complete the course. If this hypothesis held true, Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course.

The unit of diversion is a cookie, although if the student enrolls in the free trial, they are tracked by user-id from that point forward. The same user-id cannot enroll in the free trial twice. For users that do not enroll, their user-id is not tracked in the experiment, even if they were signed in when they visited the course overview page.

Experiment Design

Available Metrics

The practical significance boundary for each metric, that is, the difference that would have to be observed before that was a meaningful change for the business, is given in parentheses. All practical significance boundaries are given as absolute changes.

Any place "unique cookies" are mentioned, the uniqueness is determined by day. (That is, the same cookie visiting on different days would be counted twice.) User-ids are automatically unique since the site does not allow the same user-id to enroll twice.

- Number of cookies: That is, number of unique cookies to view the course overview page. (dmin=3000)
- Number of user-ids: That is, number of users who enroll in the free trial. (dmin=50)
- Number of clicks: That is, number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is trigger). (dmin=240)
- Click-through-probability: That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page. (dmin=0.01)
- Gross conversion: That is, number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button. (dmin= 0.01)
- Retention: That is, 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. (dmin=0.01)
- Net conversion: That is, 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. (dmin= 0.0075)

Metric Choice

Invariant Metrics

Below are the invariant metrics selected to be used in this A/B Testing experiment, followed by a justification for the classification of these metrics:

- Number of cookies: That is, number of unique cookies to view the course overview page. (dmin=3000). This metric was labeled as an invariant metric because, cookies are gathered regardless if the user chooses to click "Start free trail" or "Access course materials". The number of cookies is utilized as a unit of diversion. The number of cookies should remain significantly unchanged from the control group to the experiment group.

- Number of clicks: That is, number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is trigger). (dmin=240). This metric was labeled as an invariant metric because, number of clicks are gathered when the user decides to click "Start free trail", which is before the experiment is tested on the user. The number of clicks (clicks of the "Start free trail" button), should remain significantly unchanged from the control group to the experiment group.

- Click-through-probability: That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page. (dmin=0.01). This metric was labeled as an invariant metric because, Click-through-probability are gathered when the user clicks "Start free trial", which is before the experiment is tested on the user. The Click-through-probability, should remain significantly unchanged from the control group to the experiment group.

Evaluation Metrics

Below are the evaluation metrics selected to be used in this A/B Testing experiment, followed by a justification for the classification of these metrics:

- Gross conversion: That is, number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button. (dmin= 0.01). This metric was labeled as evaluation metric because, Gross conversion measures the effect of the experiment on the non-student user to student user measurement. This metric should decrease in the experiment, relative to the control group.

- Net conversion: That is, 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. (dmin= 0.0075). This metric was labeled as evaluation metric because, this measures the number of students to remain in the course long enough to make a payment. This metric should remain unchanged in the experiment, relative to the control group.

Unused Metrics

Below are the evaluation metrics **not** selected to be used in this A/B Testing experiment, followed by a justification for the classification of these metrics:

- Number of user-ids: That is, number of users who enroll in the free trial. (dmin=50). This metric was not used because, the Gross Conversion and Net Conversion rates reflect the number of user-ids in a much more useful manner than just the raw Number of User-Ids count.

- Retention: That is, 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. (dmin=0.01). This metric was not used because, Retention measures the probability of users who remain past 14 days and pay for enrollment. Unfortunately, if Retention is used as an evaluation metric, 4,740,020 samples must be gathered before conclusions can be drawn. For a rate of traffic of 40,000 samples per day, this will take 119 days before this metric produces results, which is too long for this A/B Test.

Measuring Standard Deviation

Baseline Values

Below is the table of baseline values for the Udacity A/B Testing Experiment.

Baseline Metric	Metric Value
Unique cookies to view page per day	40000
Unique cookies to click "Start free trial" per day	3200

Baseline Metric	Metric Value
Enrollments per day	660
Click-through-probability on "Start free trail"	0.08
Probability of enrolling, given click	0.20625
Probability of payment, given enroll	0.53
Probability of payment, given click	0.1093125

Evaluation Metric Standard Deviations

Below is a table of the analytical estimated standard deviations of the evaluation metrics used, given a sample size of 5000 cookies visiting the course overview page.

Evaluation Metric	SD
Gross Conversion	0.0202
Net Conversion	0.0156

The Gross Conversion and Net Conversion metrics are calculated with a sample size of 5000 cookies visiting the course overview page, making n = 400 for each analytical estimate. The Number of Cookies is a unit of diversion, which ensures that these calculated analytical SD estimates, are near the empirical values of SD. These measurements assume a binomial distribution for the baseline values.

Sizing

Number of Samples vs. Power

After careful consideration, the Bonferroni Correction will not be used in the analysis calculations, due to the Gross and Net Conversion metrics being correlated.

- α = 0.05
- ß = 0.2

The sample sizes needed for each evaluation metrics was calculated using this free online tool.

- Gross Conversion - Sample Size....25,835
- Net Conversion - Sample Size.....27,413

Given that these two metrics are correlated and the Click-Through-Probability for the baseline values equals 0.8:

- Pageviews (per group) = 27413 / 0.08 => 342,662.5
- Total Pageviews (both groups) = 342662.5 * 2 => 685,325

Duration vs. Exposure

The duration of the experiment by design must be at least 14 days long to collect the Net Conversion metric being measured in this experiment. Given that 40,000 pageviews occur on average everyday, and 685,325 pageviews are needed to conduct the experiment. An experiment length of 29 days was chosen, which will require 60% of total daily traffic to be exposed to the experiment. This is also a suitable duration for the experiment given that it is almost equal to 1 full month and can help temporal abnormalities.

- Experiment Length.....29 days
- Percent of Traffic Used...60%

The risk associated with this experiment is extremely low given that the data sensitivity for the metrics gathered is relatively low, the user experience is the same for existing users, and the experiment is setup in such a way to only effect new users in a positive or constructive way.

Experiment Analysis

Sanity Checks

For each metric that was labeled as an invariant metric, the 95% confidence interval for the value expected to be observed will be calculated.

The actually observed value will be listed for comparison and to conduct a Sanity Check of the invariant metrics.

Invariant Metric	Lower Bound	Upper Bound	Observed	Sanity Check
Number of cookies	0.4988	0.5012	0.5006	Passes
Number of clicks	0.4959	0.5041	0.5005	Passes
Click-through-probability	-0.0013	0.0013	0.0001	Passes

Result Analysis

Effect Size Tests

For each of the evaluation metrics, the confidence interval around the difference will be calculated without using the Bonferroni Correction. The values from these calculations will be used to determine the statistical and practical significance of each evaluation metric.

Evaluation Metric	Lower Bound	Upper Bound	Statistical Significance	Practical Significance
Gross Conversion	-0.0291	-0.0120	Yes	Yes

Evaluation Metric	Lower Bound	Upper Bound	Statistical Significance	Practical Significance
Net Conversion	-0.0116	0.0019	No	No

The Gross Conversion changes due to the effects of the experiment on the user group. This metric changing in a statistically significant way demonstrates that the Udacity Experiment did have an impact on the users.

The Net Conversion does not change due to the effects of the experiment on the user group in a statistically or practically significant way.

Sign Tests

A sign test, not using the Bonferroni Correction, for each of the evaluation metrics used in the day-by-data data will be calculated using this (free online tool)[http://graphpad.com/quickcalcs/binomial1/].

Evaluation Metric	p-value	Statistical Significance
Gross Conversion	0.0026	Yes
Net Conversion	0.6776	No

Summary

The results of the Udacity Experiment based on the selected Evaluation Metrics, showed that the Gross Conversion changed and the change was statistically significant. The Gross Conversion rate decreased within the experiment group but, the Net Conversion remained unchanged. While the Net Conversion did not change in a statistically or practically significant way, due to the confidence interval having values that were above and below zero, which is not statistically significant.

The Bonferroni Correction was not used in the analysis of this experiment due to both evaluation metrics being required to show a statistically significant change, not just one of the metrics. Utilizing the Bonferroni Correction would decrease the likelyhood of a

Type-2 Error but, due to the how the Udacity Experiment was setup the Bonferroni Correction was not utilized in this analysis.

Recommentation

Based on the statistical results of the Udacity Experiment, the change to the website is not recommended due to the Gross Conversion decreasing but, the Net Conversion remaining the same.

The changes implemented in the experiment group, led to fewer users signing up for the free-trail (Gross Conversion). This was most likely due to the user being more cautious about pursuing a Udacity course given the information that they might not have, on average, enough time to dedicate to the course.

While the number of users to remain enrolled past the 14-day mark, making their first tuition payment, remained unchanged in the experiment group. This most likely demonstrates that the caliber of users enrolling in the course did not change. Therefore, the desired effect of the Udacity Experiment was not achieved.

Follow-Up-Experiment

Outline

One idea for a follow-up-experiment would be to expose non-students to an in-depth outline of the course material for each Nanodegree. Currently, board information is given to the user before registration which leaves a lot to the imagination of the user to make the decision of whether or not to pursue a paid Nanodegree.

An outline of each course, project, and most importantly skills that will be acquired by pursing a Nanodegree. Skills like, Python, R, SQL, MongoDB, D3.js, etc. This would allow the user to obtain a very detailed insight into what program they are signing up for and if they felt that they had the necessary skill set, interest, and time to invest into the Nanodegree.

Hypothesis

If the user is exposed to an extremely detailed outline of the Nanodegree, the user Net Conversion will increase.

Metrics

- Number of cookies: That is, number of unique cookies to view the course overview page. Cookies are gathered regardless if the user chooses to click "Start free trail" or "Access course materials". The number of cookies is utilized as a unit of diversion. The number of cookies should remain significantly unchanged from the control group to the experiment group.
- Gross conversion: That is, number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button. Gross conversion measures the effect of the experiment on the non-student user to student user measurement. This metric should decrease in the experiment, relative to the control group.
- Net conversion: That is, 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. This measures the number of students to remain in the course long enough to make a payment. This metric should increase in the experiment, relative to the control group.

References

- https://docs.google.com/document/u/1/d/1aCquhIqsUApgsxQ8-SQBAigFDcfWVVohLEXcV6jWbdI/pub?embedded=True
- https://docs.google.com/document/d/16OX2KDSHI9mSCriyGIATpRGscIW2JmByMd0ITqKYvNg/edit
- https://review.udacity.com/#!/projects/4110338963/rubric
- https://drive.google.com/file/d/0ByAfiG8HpNUMakVrS0s4cGN2TjQ/view
- https://docs.google.com/spreadsheets/d/1MYNUtC47Pg8hdoCjOXaHqF-thheGpUshrFA21BAJnNc/edit#gid=0
- https://docs.google.com/spreadsheets/d/1Mu5u9GrybDdska-ljPXyBjTpdZIUev_6i7t4LRDfXM8/edit#gid=0
- http://www.evanmiller.org/ab-testing/sample-size.html
- http://graphpad.com/quickcalcs/binomial1/