**Experiment - Udacity**

*Introduction*

At the time of this experiment, Udacity courses currently have two options on the course overview 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.

*Metric Choice*

We first need to choose invariant metrics. We will use these metrics to make ourselves sure that control and experiment group division has been performed unbaised.

**Invariant Metrics**:

* **Number of cookies**: That is, number of unique cookies to view the course overview page. (dmin=3000)
* **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)

Evaluation metrics will be used for significance tests which will show us whether the change made in the website has a real effect or not.

**Evaluation 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*)
* **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*)

|  |  |  |
| --- | --- | --- |
|  | Values | Standard Devs. |
| Unique cookies to view course overview page per day: | 5000 |  |
| Unique cookies to click "Start free trial" per day: | 400 |  |
| Enrollments per day: | 82.5 |  |
| Click-through-probability on "Start free trial": | 0.08 |  |
| Probability of enrolling, given click: | 0.20625 | 0.0202 |
| Probability of payment, given enroll: | 0.53 | 0.0549 |
| Probability of payment, given click | 0,1093125 | 0.0156 |

We do not prefer to use Bonferroni Correction since three evaluation metrics we use look quite correlated, by which Bonferroni method will cause a very conservative solution.

*Power, Sample Size Calculation and Duration Considerations*

Given the significance levels of each evaluation metric we’ll calculate the necessary amount of samples for confidence interval of 95% and power of 80% using an online calculator. Necessary amount of samples are given below:

**Gross Conversion**

* Baseline Conversion: 20.625%
* Minimum Detectable Effect: 1%
* Sample Size = 25,835 enrollments/group
* Total sample size = 51,670 enrollments (2 groups)
* Clicks/Pageview: 3200/40000 = 0.08 clicks/pageview
* Pageviews Required = 51,670 / 0.08 = 645,875

**Retention**

* Baseline Conversion: 53%
* Minimum Detectable Effect: 1%
* Sample size = 39,155 enrollments/group
* Total sample size = 78,230 enrollments (2 groups)
* Enrollments/pageview: 660/40000 = 0.0165 enrollments/pageview
* Pageviews Required = 78,230/0.0165 = 4,741,212

**Net Conversion**

* Baseline Conversion: 10.9313%
* Minimum Detectable Effect: 0.75%
* Sample size = 27,413 enrollments/group
* Total sample size = 54,826 enrollments (2 groups)
* Enrollments/pageview: 0.08 clicks/pageview
* Pageviews = 54,826 / 0.08 = 685,325

The largest amount should be taken to make sure all of the metrics are at safe zone, which is for Retention and 4,741,212 pageviews. However, it is known that there are 40,000 pageviews to Udacity website a day, and this means that the experiment will take about 119 days, which can be considered too long for an A/B test.

We should rather drop Retention and stick to Gross and Net Conversion metrics. In order to keep the duration low enough, 100% diversion could be chosen that ends up with an 18 day (685,325 / 40,000) experiment if and only if there are no other experiments conducting at the same time.

*Sanity Checks*

In this part, we are going to use invariant metrics to perform the sanity checks. For the first two of the invariant metrics, we’ll calculate if the division to control and experiment groups had been done appropriately (with probability of 0.5 or not). All intervals will be calculated for 95% confidence level.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Control Group | Experiment Group | Lower Bound | Upper Bound | Observed | Passed? |
| Number of cookies | 345543 | 344660 | 0.4988 | 0.5012 | 0.5006 | Yes |
| Number of clicks | 28378 | 28325 | 0.4959 | 0.5041 | 0.5005 | Yes |

First two count metrics passes the sanity check. Now is the time to calculate a confidence integral for third invariant metric.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Expected | Observed | Lower Bound | Upper Bound | Passed? |
| Click-through probability | 0.0821 | 0.0822 | 0.0761 | 0.0883 | Yes |

This metric also passes the sanity check. So far, everything looks fine to proceed for the result analysis.

*Result Analysis*

95% rate has been chosen as the confidence interval for the difference between the experiment and control group for evaluation metrics. The result is statistically significant only when the 95% confidence interval does not include zero. The result can be considered practically significant if aforementioned interval does not include the practical significance level, dmin.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **dmin** | **Observed Difference** | **Lower Bound** | **Upper Bound** |
| Gross Conversion | 0.01 | -0.0205 | -0.0291 | -0.0120 |
| Net Conversion | 0.0075 | -0.0048 | -0.0116 | 0.0019 |

### From the table above, it is seen that Gross Conversion metric results are both statistically and practically significant. However on the other hand, Net Conversion metric results are neither statistically nor practically significant.

*Summary and Recommendation*

In this experiment, an A/B test is performed to measure the effect of and additional screen at Udacity, which asks for the hours devoted a week to the students and gives an advice that if the amount of hours given by the user is less than 5 hours, it’d be better to access the course content without enrollment. Sanity checks have been done using three metrics to make sure that control and experiment groups are created unbaised. Two evaluation metrics were used which are Gross Conversion (enrollment per cookie) and Net Conversion (payments per cookie). In the end, it is found that result for Gross Conversion is practically significant, but result for Net Conversion is not even statistically significant.

This result means that, even though number of enrolled students has been decreased significantly, number of payments did not. This was the exact aim of the experiment to show, therefore I recommend to launch the change at this point.