****Project 7 Report: A/B Test****

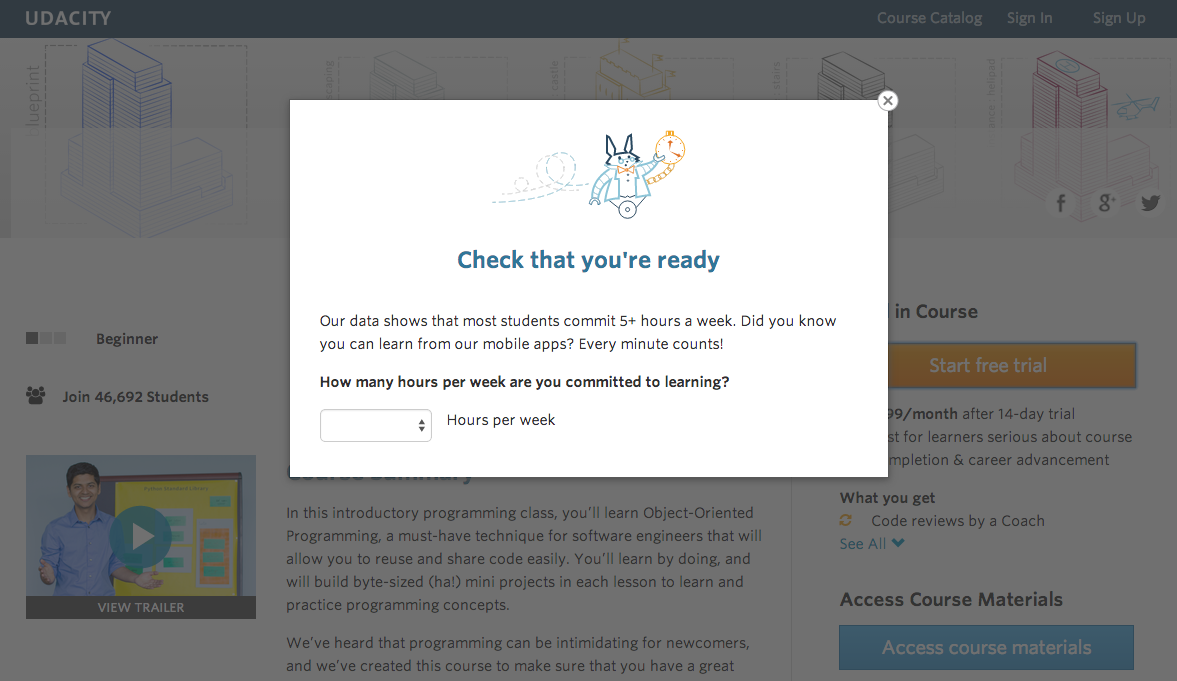
## by Tang Shuran

## Experiment Description

For control group, Udacity courses currently have two options on the home page: "start free trial", and "access course materials". The "start free trial" option lead the users to a credit card information input page. The users will get a 14 days free trial after they filled the form. If the user did not cancel the subscription in 14 days, they will be charged automatically after the trial finish.

"access course materials" option lead the users directly to the videos and quizzes, but they will not receive coaching support or a verified certificate, and they will not submit their final project for feedback.

For experiment group, most procedures are in same way. But a change is made when the student clicked "start free trial" option. Instead of leaded to the credit card information form, they will be asked that, how long are they willing to spend their time on the Udacity course. 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.



The purpose of this change is to reduce 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.

# Experiment Design

## Metric Choice

***Invariant Metrics:***number of cookies, number of clicks, click-through-probability

**Number of Cookies:**  The number of unique cookies to visit the course overview page. It is the unit of diversion in this A/B test. Therefore, it is expected to be the same between the control and experiment group.

**Number of Clicks:** The number of users to click the free trial button. Because the change we made in experiment group do not appear until the user click the button, it doesn’t impact the clicking button decision of the users.

**Click-through-probability:** Unique cookies to click the "start free trial" button per unique cookies to view the course overview page. Because it is a number that clicks devided by cookies, it will not change if the number of clicks and cookies doesn’t change.

***Evaluation Metrics:***gross conversion, retention, net conversion

Evaluation metrics were chosen to measure the difference between control group and experiment group. In this context, the goal of our choice is to minimize students’ frustration and provide most satisfaction with effective usage of limited coaching resources. In order to achieve this, we expect to observe following changes:

**Gross conversion:** 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. For the launch of the change, the gross conversion is indicated to decrease. Because the potential frustrated students may be filtered out.

**Retention:** number of user-ids to remain enrolled past the 14-day boundary divided by number of user-ids to complete checkout. If our hypothesis stands, our change will not significantly reduce the number of students who remain enrolled past the 14 days. But the experiment change will lead to less users for check out. Therefore, a slight increase of retention is expected.

**Net Conversion:**number of user-ids to remain enrolled past the 14-day boundary divided by the number of unique cookies to click the "Start free trial" button. Because clearer expectations are set for students upfront, more students may remain enrolled past the 14 days. Therefore, a slight increase of net conversion is expected.

***Unused Metrics:*** Number of user-ids

**Number of user-ids:** The number of users to enroll in the free trial. it can be used as evaluation metric because it would track the first part of the hypothesis, but we don’t use it this time because it is not normalized.

## Measuring Standard Deviation

In the experiment, we predict that we will need approximately 5,000 cookies per day in each  
sample group. Corresponding results are shown below:

|  |  |  |
| --- | --- | --- |
|  | baseline | sample |
| Unique cookies to view page per day: | 40000 | 5000 |
| Unique cookies to click "Start free trial" per day: | 3200 | 400 |
| Enrollments per day: | 660 | 82.5 |
| Click-through-probability on "Start free trial": | 0.08 | 0.08 |
| Probability of enrolling, given click: | 0.20625 | 0.20625 |
| Probability of payment, given enroll: | 0.53 | 0.53 |
| Probability of payment, given click | 0.109313 | 0.109313 |

The Analytical estimate of standard deviation of these normally distributed rates is:

Then we can calculate the standard deviation of evaluation metrics.

|  |  |
| --- | --- |
| **Evaluation Metric** | **Standard Deviation** |
| Gross Conversion | 0.0202 |
| Retention | 0.0549 |
| Net Conversion | 0.0156 |

The analytical estimate of standard deviation is close to empirically determined standard deviation if the unit of diversion is equal to the unit of analysis. Therefore, for Gross Conversion and Net Conversion, we can use analytical estimate. If we need to use Retention, we should calculate its empirical variability.

## Sizing

### Number of Samples vs. Power

In order to know the exact number of pageviews required for our experiment, we should find the sample size that we will need for each evaluation metric first. Then we can calculate the required pageviews from the sample size.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Evaluation Metric** | **Baseline Conversion** | **Minimum Detectable Effect** | **alpha** | **beta** | **sample size** | **total sample size** | **pageviews** |
| Gross Conversion | 20.625% | 1% | 5% | 20% | 25835 | 51670 | 645875 |
| Retention | 53% | 1% | 5% | 20% | 39155 | 78230 | 4741212 |
| Net Conversion | 10.9313% | 0.75% | 5% | 20% | 27413 | 54826 | 685325 |

### 

In conclusion, the largest pageviews number is our limiting factor, so we need 4741212 pageviews for our experiment.

### Duration vs. Exposure

Because the experiment change is just a mild reminder for the users, it can’t result into physical or mental damage. Additionally, no sensitive information can be collected during these procedures. Therefore, 100% of the traffic can be divert into the experiment.

If we divert 100% of the traffic into the experiment, it takes us 119 days to collect enough data for analysis. It is too long for our experiment. Hence, i choose to delete the retention from the evaluation metric list. As a result, the required pageviews number decrease to 685325. Consequently, we need only 18 days to get enough data. That would be a reasonable duration.

# Experiment Analysis

## Sanity Checks

1. **Number of Cookies**

The expected number of cookies in experiment group should be equal to control group. So the expected rate of diversion should be 0.5. So we will construct a 95% confidence interval around 0.5 for comparing with observed rate.

P=0.5

α=0.05

Z=1.96

=0.006018

ME=1.96 \* 0.00601841 = 0.0011796

CI = [0.4988, 0.5012]

Observed rate = 345543/(345543+344660)= 0.50064

Observed rate is within the confidence interval.

1. **Number of clicks**

The expected number of clicks in experiment group should be equal to control group. So the expected rate of diversion should be 0.5. So we will construct a 95% confidence interval around 0.5 for comparing with observed rate.

P=0.5

α=0.05

Z=1.96

= 0.002099747

ME=1.96 \* 0.002099747= 0.004115504

CI = [0.4959, 0.5041]

Observed rate = 28378/(28378+28325)= 0.500467347

Observed rate is within the confidence interval.

1. **Click-through-probability**

The expected Click-through-probability in experiment group should be equal to control group. So we will construct a 95% confidence interval around 0.5 for comparing with observed rate.

P=28378/345543= 0.082125814

α=0.05

Z=1.96

= 0.00467

ME=1.96 \* 0.00467= 0.00091532

CI = [0.0812, 0.0830]

Observed rate = 28325 / 344660 = 0.082182

Observed rate is within the confidence interval.

**In conclusion, all the 3 invariant metrics pass the sanity check.**

## Result Analysis

### Effect Size Tests

For each evaluation metric, we need to find the degree of the difference between the control and experiment group. We firstly construct a 95% confidence interval for each metric. Then we compare the minimum detectable difference dmin. with it to decide the significance.

1. Gross Conversion

α= 0.05

Z = 1.96

Rc=0.21887468

Rx=0.19831981

d̂= -0.020554874

Varc==9.88657605 \* 10-6

Varx==9.211417482 \* 10-6

Vard̂= Varc + Varx= 1.909799252 \* 10-5

σd̂== 4.370125-3

ME = 8.5652 \* 10*-*3

CI= [-0.0291, -0.0120]

For Gross Conversion, the minimum detectable effect equals to 0.01. The confidence interval in our experiment is outside this range. The observed difference is also outside the confidence interval.

Therefore, we can decide that, Gross conversion change is both statistically and practically significant.

1. Net Conversion

α= 0.05

Z = 1.96

Rc= 0.1175620193

Rx= 0.1126882966

d̂= -0.004873723

Varc==5.999027983 \* 10-6

Varx==5.793142782 \* 10-6

Vard̂= Varc + Varx= 1.179217076 \* 10-5

σd̂== 3.43397029\*10-3

ME = 6.7228 \* 10*-*3

CI= [-0.0116, 0.0018]

For Gross Conversion, the minimum detectable difference equals to 0.0075. The confidence interval in our experiment is not outside this range. The observed difference is within the confidence interval.

Therefore, we can decide that, Gross conversion change is neither statistically nor practically significant.

### Sign Tests

|  |  |  |
| --- | --- | --- |
| Metric | p-value for sign test | Statistically Significant? |
| Gross Conversion | 0.0026 | Yes |
| Net Conversion | 0.6776 | No |

### 

### Summary

In this experiment, we divert the traffic to Udacity website into 2 groups. For experiment group, every unique cookie will be asked the time we are willing to spend on udacity course before the enrollment. For comparison with the control group, Gross conversion and Net Conversion are used as the evaluation metrics.

The result shows that, the gross conversion change is both statistically and practically significant. But the net conversion change is neither statistically nor practically significant.

The Bonferroni correction is used for avoiding false positive error when we try to find whether someone of the multi evaluation metrics meets our specification. However, in our case, we decide to launch only if all the evaluation metrics meet our specification. In this situation, as the number of metrics increases, the risk of type two error increase but not type one.

Because the Bonferroni correction is used to avoid type one error, it is not appropriate for our experiment.

## Recommendation

The aim of this experiment is to find that, whether our process change can filter out the students who have no enough time for Udacity courses and improve the overall students experience with limited couch resource.

The experiment result illustrates that the Gross Conversion significantly decrease. It means that, our change really filtered out some students. But the confidence interval of the net conversion does include the negative of the practical significance boundary, it is possible that net conversion drops due to our experiment change, which doesn’t meet second part of our hypothesis.

Based upon this finding, I recommend that we should not launch this change. It may harm the interests of Udacity.

# 

# Follow-Up Experiment

For follow-up experiment, we can add a reminder message window for the students who have enrolled but didn’t commit enough time in the first week, and ask them whether they are willing to spend more time next week. If the answer is yes, the user will be led directly to the course learning page. But if it is no, the user will be asked whether they are willing to drop out.

This experiment is based on the hypothesis that, some students would not actually commit as much time as they expected, and we should remind them of this fact and give them new choice.

This new option page will improve the experience of students by filtering out the potential frustrated students in the future, and it will also reduce the work load of Udacity couches.

**Null Hypothesis**: The new reminder window can‘t filter out the students who would not continue after 14 days trial from the students spending not enough time in the first week.

**Unit of Diversion:**  user-id, because this experiment is based on the time each student spends on the course. The time is counted based on each unique user-id.

**Invariant Metrics:** The invariant metric should be the number of user-id. Because each user is tracked along the experiment for a week, then cookies are not suitable in this situation. **Evaluation Metrics:** Evaluation metric should be the number of early cancelling students. A slight increase for the early cancelling students is expected, because some users are reminded that their time plans don’t fit the Udacity requirements.