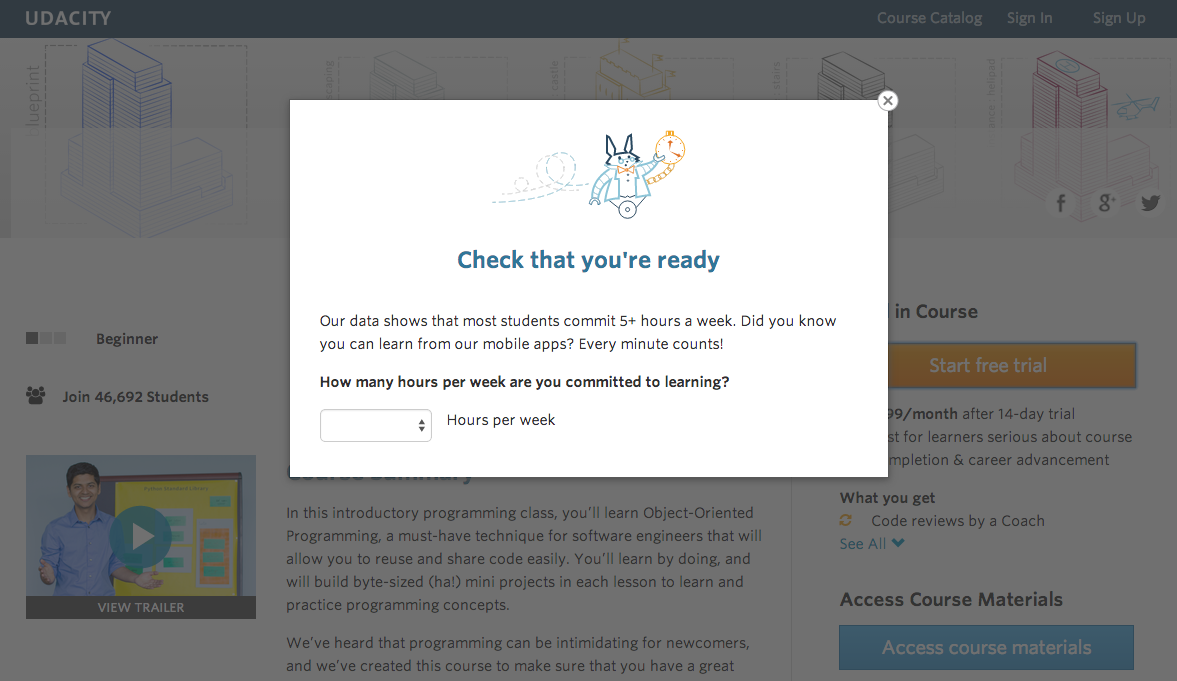
****Project 7 Report : A/B Test****

## 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. It is indicated to decrease, because the experiment change will lead some users, who expected to spend less than 5 hours per week, to free material other than enrollment while the clicks number do not change.

**Retention:** number of user-ids to remain enrolled past the 14-day boundary divided by number of user-ids to complete checkout. Due to the decrease number of user-ids to complete checkout, users who intend to spend not enough time are filtered out, therefore the proportion of potential retainer in all enrollments will increase.

**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. A slight decrease is expected, because retainers number, affected by the change, may decrease. At the same time, clicks number remain the same.

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

**Number of user-ids:** The number of users to enroll in the free trial. It is not suitable for both invariant metric and evaluation metric because User-ids are only tracked after enrollment.

#For each metric, explain both why you did or did not use it as an invariant metric and why you did or did not use it as an evaluation metric. Also, state what results you will look for in your evaluation metrics in order to launch the experiment.

## 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 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 |

#For each of your evaluation metrics, indicate whether you think the analytic estimate would be comparable to the empirical variability, or whether you expect them to be different (in which case it might be worth doing an empirical estimate if there is time). Briefly give your reasoning in each case.

## 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 |

### 

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

### Duration vs. Exposure

Indicate what fraction of traffic you would divert to this experiment and, given this, how many days you would need to run the experiment. (These should be the answers from the "Choosing Duration and Exposure" quiz.)

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?

# Experiment Analysis

## Sanity Checks

For each of your invariant metrics, give the 95% confidence interval for the value you expect to observe, the actual observed value, and whether the metric passes your sanity check. (These should be the answers from the "Sanity Checks" quiz.)

For any sanity check that did not pass, explain your best guess as to what went wrong based on the day-by-day data. **Do not proceed to the rest of the analysis unless all sanity checks pass.**

## Result Analysis

### Effect Size Tests

For each of your evaluation metrics, give a 95% confidence interval around the difference between the experiment and control groups. Indicate whether each metric is statistically and practically significant. (These should be the answers from the "Effect Size Tests" quiz.)

### Sign Tests

For each of your evaluation metrics, do a sign test using the day-by-day data, and report the p-value of the sign test and whether the result is statistically significant. (These should be the answers from the "Sign Tests" quiz.)

### Summary

State whether you used the Bonferroni correction, and explain why or why not. If there are any discrepancies between the effect size hypothesis tests and the sign tests, describe the discrepancy and why you think it arose.

## Recommendation

Make a recommendation and briefly describe your reasoning.

# Follow-Up Experiment

Give a high-level description of the follow up experiment you would run, what your hypothesis would be, what metrics you would want to measure, what your unit of diversion would be, and your reasoning for these choices.