# Template Format

This template can be used to organize your answers to the final project. Items that should be copied from your answers to the quizzes should be given in blue.

# Experiment Design

## Metric Choice

List which metrics you will use as invariant metrics and evaluation metrics here. (These should be the same metrics you chose in the "Choosing Invariant Metrics" and "Choosing Evaluation Metrics" quizzes.)

**Invariant metrics:**

**Number of cookies:** That is, number of unique cookies to view the course overview page.

**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).

**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.

**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.

**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.

**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.

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.

Invariant metrics are metrics that should not be affected by the experiment. This metrics have a 50/50 chance of getting assigned to control or experiment group and the difference between the two groups have to indicate that, otherwise there is something wrong. Evaluation metrics are metrics that are expected to present or not a difference between the control and experiment group.

**Number of cookies** should not change between control and experimental group since this metrics is not affected by the experiment, it indicates the number of cookies to view the overview page. The change belonging to the experiment is not going to affect the ability of the user to get to this page because of that it is an invariant metric. This metric is not an evaluation metric because it should not get affected by the experiment and therefore can not be used to evaluate the experiment.

**Number of user-ids** is not a invariant metric because we expect the number of users to be different between the two groups. It could be used as an evaluation metric because it can tell if actually there is a difference between the number of users that enrolled, the problem of doing so is that this information is not robust enough. This metric can both represent users that started the free trial and surpassed the 14 days mark and users that started the free trial and did not surpassed the 14 days mark. Also other metrics that are more robust like gross conversion and net conversion can give us the same information that this variable would do.
**Number of clicks** is an invariant metric because it computes the number of cookies that clicked in the start-free-trial button, this metric should not be affected by the experiment because all cookies that are in the course overview page can have access to the button independent if the cookie belongs to control or experimental group. This metric is not an evaluation metric because it does not get affected by the experiment.
**Gross conversion** represents the proportion of users that were in the overview page and clicked in the start-free-trial button by the number of users who enrolled in the free trial. It is expected from this metric to present a difference between the control and experimental groups because it is computed with data that is directly affected by the experiment, that is the number of users that enrolled in the free trial. Since it is expected for this metric to be affected, it can not be a invariant metric because it may present a difference between control and experiment groups.
**Retention** is the proportion of enrolled users that passed the 14 days mark divided by the number of enrolled users. It is expected that the proportion is going to be higher in the experimental group because the number of users that are going to be frustrated for not having enough time to do the course are going to be less than in the control group, what leads to a higher proportion of users that are going to pay. Since it is expected a change for this variable it can not be used as a invariant metric.
**Net conversion** is the proportion of users that surpassed the 14 days mark divided by the number of cookies that clicked the start-free-trial button. It is expected for the experiment group to have a higher proportion since this group is expected to have a larger number of users that are going to surpass the mark. The number of cookies that clicked the start-free-trial button should stay the same since this is a invariant metric. Net conversion can not be an invariant metric because it is expect for its value to change between control and experimental groups.

We expect from the experiment that the evaluation metrics are going to demonstrate that the experiment yields what it was expect from it, that is to increase the rate of student that enroll in the free trial and surpass the 14 days mark, meaning that the number of users frustrated is smaller.
The expected behave for gross conversion is to show that the number of users enrolled in the experiment group is smaller than in the control group since we want to reduce the number of users that do not have time to dedicate to the course.
Retention only takes in consideration users that enrolled in the program, so its expected behavior is a larger ratio for the experiment group.

Now for net conversion, we expect a reduction in the ratio of users that surpassed the 14 days mark by the number of cookies to click the start free trial button. Although we expect a reduction, it has to be in the limits that the company agrees, that is inside the practical limits.

## Measuring Standard Deviation

List the standard deviation of each of your evaluation metrics. (These should be the answers from the "Calculating standard deviation" quiz.)

**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 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.

The decision whether the analytic estimate would be compared to the empirical variability is going to be based purely on the unit of analysis and the unit of diversion. For each evaluation metric the unit of diversion is cookies. The test randomly places each cookie inside one of the groups, and that cookie remains in that group forever.

For gross conversion and net conversion the unit of analysis is also cookies. That mean we do not have to worry about creating an empirical estimate.

Regarding retention, the variable is the number of user-ids to remain enrolled past the 14-day boundary divided by number of user-ids to complete checkout. That means the unit of analysis is user-ids, that is different of cookies. In this case, the analytic estimate can differ from empirical estimates. For instance, suppose that we have an user that studies on desktop and has the browser configured to clean the cookies every time it is closed. For this user we are going to have multiple cookies assigned to it, and that is not what we want. In this case we will probably want to run empirical estimates to get the true effect this has on our metric.

## Sizing

### Number of Samples vs. Power

Indicate whether you will use the Bonferroni correction during your analysis phase, and give the number of pageviews you will need to power you experiment appropriately. (These should be the answers from the "Calculating Number of Pageviews" quiz.)

I am not going to be using Bonferroni corrections during my analysis.

In this stage I am only going to be using gross conversion and net conversion as evaluation metrics. That is because retention needs a large number of pageviews that are not possible to be completed in a short period of time with the traffic we have.

I am going to be needing 685325 pageviews.

### 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.)
Fraction of the traffic diverted 1.
Number of days that the experiment would need to run 18.

Give your reasoning for the fraction you chose to divert. How risky do you think this experiment would be for Udacity?
Since the experiment is not going to cause harm to anyone and is not collecting any kind of information from the users, it is low risk for Udacity. It is a small change and the chance of bug are minimal. Also taking in consideration that the client wants the experiment to take less than 30 days, I did convert 100% of the traffic to the experiment.

# 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.)

| Variable | Lower bound | Upper bound | Observed | Passes |
|---|---|---|---|---|
| Number of cookies | 0.4988 | 0.5012 | 0.5006 | Yes |
| Number of clicks | 0.4959 | 0.5041 | 0.5005 | Yes |
| Click-through-probability | -0.0013 | 0.0013 | 0.0001 | Yes |

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.**
All invariant metrics passed the sanity check.

## 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.)

| Variable | Lower bound | Upper bound | Statistically significant | Practically significant |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| Gross conversion | -0.0291 | -0.0120 | Yes | Yes |
| Net conversion | -0.0116 | 0.0018 | No | No |

### 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.)

| Variable | P value | Statistically significant |
|---|---|---|
| Gross conversion | 0.0026 | Yes |
| Net conversion | 0.6776 | No |

### 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.

If multiple hypotheses are tested, the chance of a rare event increases, and therefore, the likelihood of incorrectly rejecting a null hypothesis increases. The Bonferroni correction compensates for that increase by testing each individual hypothesis at a significance level of alpha/m, where m is the number of test.
The probability of observing at least one significant result due to chance remains below the desired significance level alpha, but can be much more conservative in rejecting the null hypothesis.
In the case of multiple metrics, we can tell with metric to use just on knowing whether or not we have the ALL or ANY case for metrics and to determine that we need to understand the experiment and what each variable is measuring.
For this experiment I would say that we have the ALL case scenario for this metrics. That means we are only going to launch the experiment if gross conversion and net conversion mets its conditions.
In an ANY case scenario, one false positive is potentially more harmful because even if it affects one variable, the experiment would still be launched. That is way in this cases it makes more sense to use Bonferroni correction being more conservative. In an ALL case scenario, one false positive has less effect since all variables have to meet the requirements, that is way we do not need to be so conservative.

There is no discrepancy in the results, gross conversion did pass on both effect size test and also sign test. Net conversion did not pass on both effect size test and neither on sign test.

### Recommendation

Make a recommendation and briefly describe your reasoning.

My recommendation is to not launch the experiment.
Gross conversion did passed the test but net conversion did not. That means we had a statistically difference in the number of users that did enroll in the course between the two groups but we do not have a statistically significant neither a practically significant difference between the number of users that surpassed the 14 days mark.
Net conversion lower bound is negative and higher than the practical significance of 0.0075 and the upper bound is below the positive practical significance threshold. That means Udacity may be losing more money than is expected.

## 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.

My suggestion for the follow-up experiment would be an improvement in the overview page. The reason for this is that in the previous experiment we did see that removing the frustrated users from the free trial may cause a negative impact in Udacity's revenue. What we are trying this time is to not remove this users, but remove the users that were in the free trial only to get additional information that they did not have from the overview page.
In the overview page I would cover all the course material, tutors and examples of what the user is going to learn in the course.
My hypothesis is that users that are only starting the free trial to get more information about the course are impacting the relation of users that start the free trial and surpass the 14 days mark.
The metrics that I would want to measure are:
**Invariant metrics:**
Number of cookies: That is, number of unique cookies to view the course overview page.
**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.
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.
My unit of diversion is going to be cookie. Cookies are going to be our unit of diversion because all of our metrics have cookies as unit of analysis.