Experiment Overview (provided by Udacity)

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.

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

Metric Choice

There are two types of metrics in an experiment: invariant metrics and evaluation metrics. Invariant metrics are the ones you expect to not change in the experiment. It is used as a “sanity check” to ensure the validity of the experiment, to make sure the setup is not inherently wrong. On the other hand, evaluation metrics are the ones you expect to change and are related to the business goals we want to achieve with the experiment results. For each metric, I have defined a practical significance, d_{min}, which indicates the minimum change that is considered significant to the business.

Invariant metrics

- Number of cookies: Number of unique cookies to view the course overview page. If this changes, that means there is a change in traffic before the Free Trial Screener feature is triggered, which would be problematic.
- Number of clicks: Number of unique cookies to click the “Start free trial” button. This happens before the feature is triggered, so it should not change.
- Click-through-probability: Number of unique cookies to click the “Start free trial” button divided by the number of unique cookies to view the course overview page. Following the same logic, this value should not change.

Evaluation metrics

- Gross conversion: Number of user-ids to complete enrollment in the free trial divided by the number of unique cookies to click the "Start free trial" button. Changes in gross conversion will show how the feature affected the number of users who enrolled in the free trial even after answering the question about their available time to devote to the course. I would look for at least a 0.01 change in gross conversion.
- Retention: Number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of user-ids to complete checkout/enroll. We would look for at least a 0.01 change in retention.
- Net conversion: 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. We would want at least a 0.0075 change in net conversion.

Unused metric:

- Number of user-ids: Number of users who enroll in the free trial. This cannot be used as an invariant metric because it is tracked after enrolling in the free trial. So there would not be an equal distribution of number of user-ids between the control and the experiment groups. It cannot be used as an evaluation metric either since it does not directly represent the success of the feature. It could be used to track the number of users who remained enrolled past the 14-day boundary, but the other metrics are better not absolute values but proportions with respect to enrollments/clicks.

Measuring Standard Deviation

Baseline Values

Unique cookies to view course overview page per day:	40000
Unique cookies to click "Start free trial" per day:	3200
Enrollments per day:	660
Click-through-probability on "Start free trial":	0.08
Probability of enrolling, given click:	0.20625
Probability of payment, given enroll:	0.53
Probability of payment, given click	0.1093125

Analytically Estimated Standard Deviations

Evaluation Metric	Standard Deviation
Gross Conversion	0.0202
Retention	0.0549
Net Conversion	0.0156

- Gross conversion: The unit of diversion and the unit of analysis are both cookies. Therefore, in this case, the analytical estimate of standard deviation is sufficient.
- Retention: The unit of diversion is cookies but the unit of analysis (denominator of metric) is enrollments. In this case, the analytical value would underestimate the empirical value and we would want an empirical estimate of variability if there is time.
- Net conversion: The unit of diversion and the unit of analysis are both cookies and the analytical estimate is sufficient.

Detailed calculations can be found in the Jupyter notebook.

Sizing

Number of Samples vs. Power

I will not use Bonferroni correction because we are only testing two metrics in this experiment, so multiply hypothesis testing is not really an issue. Furthermore, gross conversion and net conversion are likely to be correlated, which means Bonferonni correction would be too conservative of an approach to correct this anyways.

In this experiment, alpha is defined as 0.05 and beta as 0.2. The following number of pageviews are required to power the experiment appropriately:

Evaluation Metric	Required Pageviews
Gross Conversion	645875
Retention	4737818
Net Conversion	685325

For retention, the required number of pageviews is over 4 million. From the baseline values, we know that Udacity gets about 40,000 pageviews per day. This would make the duration of the experiment too long, so I decided to drop this metric.

Out of the remaining two metrics, net conversion requires 685,325 views, which should be the number of samples for this experiment.

Duration vs. Exposure

Assuming that 75% of Udacity's daily traffic is diverted to this experiment, it would take 23 days to run this experiment. I chose to divert 75% because Udacity may want to run other experiments concurrently. Choosing to divert all of the traffic may be risky in that it could cause confusion among students or lead to lower conversion.
Again, detailed calculations can be found in the Jupyter notebook.

Experiment Analysis

See Jupyter notebook for detailed calculations.

Sanity Checks

Invariant Metric	95% Confidence Interval	Observed Value	Sanity Check Pass?
Number of cookies	[0.4988,0.5012]	0.5006	Yes
Number of clicks on "Start free trial"	[0.4959,0.5041]	0.5005	Yes
Click-through-probability on "Start free trial"	[-0.0013,0.0013]	0.0001	Yes

All of the invariant metrics pass the sanity check, so it is fine to move forward with the analysis.

Result Analysis

Effect Size Tests

Evaluation Metric	95% Confidence Interval	Statistical Significance?	Practical Significance?
Gross Conversion	[-0.0291,-0.012]	Yes	Yes
Net Conversion	[-0.0116,0.0019]	No	No

The change in gross conversion from the Free Trial Screener feature is both statistically and practically significant. On the other hand, the change in net converison is not significant.

Sign Tests

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

The sign tests support the results from effect size tests. The change in gross conversion was significant, but the change in net conversion wasn't.

Recommendation

The goal of the Free Trial Screener feature was to see whether filtering students based on time commitment could improve student experience and allow coaches to focus on students who are likely to complete the course. If the feature achieved this without significantly reducing the number of students who continue past the free trial, it would be successful.

Based on the results above, I recommend to not launch the feature and pursue other experiments instead. It has caused a significant change in gross conversion, which represents the probability of enrolling per click, but not in net conversion, which is more closely related to Udacity's business goals. This means that there was a decrease in enrollment, but no significant changes in students staying past the 14 days boundary to pay for the course.

Follow-Up Experiment

Aside from student frustration with unexpected amount of time commitment for a course, other factors, such as the unexpected difficulty of the course, could affect early cancellation. If a student lacks the required skills to be able to excel in the course, it could lead to frustration and early cancellation. In order to test this idea, I suggest a follow-up experiment in which the Free Trial Screener feature is modified to ask students whether they have the specific required skillsets and knowledge before enrolling in the course. The hypothesis would be that asking students about prerequisites filters out students who are less likely to complete the course, improving overall student satisfaction and leaving enough coaching resources for those who do complete the course. The metrics and the unit of diversion would stay the same as before. If we observe a significant (both statistically and practically) decrease in gross conversion and a significant increase in net conversion, we would conclude that this feature is successful.